Molecular characterization of biofilm production and whole genome sequencing of selected *Campylobacter concisus* oral and clinical strains

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Mohsina Huq
BSc; MSc (Microbiology); MS (Biotech)

School of Applied Sciences
College of Science Engineering and Health
RMIT University

September, 2016
DECLARATION

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program; any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed.

Mohsina Huq

30/09/2016
ACKNOWLEDGEMENTS

I would like to express my sincere appreciation and gratitude to my senior supervisor Dr. Taghrid Istivan for her invaluable comments and comprehensive feedback on my work. I am grateful to her for her patient support, kind assistance, consistent encouragement and great guidance throughout the entire period of my research. My sincere gratitude goes to my associate supervisors Professor Peter Smooker and Dr. Thi Thu Hao Van for their continuous encouragement and guidance throughout this research program.

I would like to thank the Department of Education and Training, Australian Government for awarding me with the Endeavour Postgraduate Scholarship. Without their funding, this work would not have been possible. My sincere gratitude goes to the Dean of Research and Innovation, Professor Andrew Smith for granting me the top-up scholarship which was a great help to complete this research study. I am grateful to Dr. Anna Walduck for her scientific editing and proofreading of my thesis. Thanks for Dr. Lisa Dias; I really appreciate your guidance, and the time you spent with me to make my PhD study go well.

I would like to thank the past and current students in Dr. Istivan’s laboratory, especially Dr. Eltaher Elshagmani and Dr. Khaled Allemailem for their continuous assistance and encouragement. I would like to express my sincere thanks to all members in the School of Science (Bundoora) for their friendship and assistance, especially Layla Alhasan, Abdouslam Alsharif, Alaa Karkashan, Mohamed Said, Natalie Kikidopoulos, Thomas Mclean, Alekhya Penumarthi, Vishal Mistry, Danka Jovetic, Kim Stevenson Huan, Yong Yap, Gor Mian Chee and Ben Vazina and to all staff members of Biotechnology and Environmental Biology for their friendship, assistance, and company.
Special thanks to Dr. Eric Adetutu and Vijay Bhaskarla for helping me with the next generation sequencing. Sincere gratitude to my associate supervisor Dr. Van for sequencing two strains in the Illumina Miseq. Special thanks to Professor Rob Moore for performing the de novo assembly of one genome and guidance in my research. I would like to acknowledge the help of Dr. Chaitali Dekiwadia and all members of RMIT Microscopy and Microanalysis Facility for letting me use the electron microscope.

I would like to acknowledge my parents, Mozammel Hoque and Ferdous Akhter, my three brothers, and my two sisters for their love, guidance and support. Their inspiration and encouragement helped me to get through the hardest times.

I would like to express my sincere appreciation to my husband Monir and especially my son Mahrus, for their love, support and patience over many years throughout my current and previous studies. I hope that this thesis will assure them that one of their dreams in addition to mine has finally come true.

Lastly, I owe everything to the Almighty God (Allah) without whose grace and guidance, any of this would be possible.
Table of contents

DECLARATION .................................................................................................................. I
ACKNOWLEDGEMENTS ..................................................................................................... II
Table of contents ................................................................................................................ IV
List of Figures ..................................................................................................................... X
List of Tables ........................................................................................................................ XIV
Summary .............................................................................................................................. XV
Publications .......................................................................................................................... XIX
Abbreviations ...................................................................................................................... XX

Chapter One : General Introduction .................................................................................. 1
  1.1 The genus Campylobacter .............................................................................................. 1
     1.1.1 History and association with diseases ................................................................. 3
  1.2 Campylobacter concisus: the emerging pathogen ............................................................. 5
     1.2.1 C. concisus in the human oral cavity ................................................................. 5
     1.2.2 C. concisus in acute gastroenteritis ..................................................................... 8
     1.2.3 C. concisus in chronic gastrointestinal diseases .................................................. 11
     1.2.4 C. concisus in immunocompromised patients .................................................... 15
     1.2.5 C. concisus in other parts of the human body .................................................... 15
  1.3 Identification of C. concisus .......................................................................................... 17
     1.3.1 Identification of C. concisus by cultural and biochemical properties ............... 17
     1.3.2 Identification of C. concisus by molecular methods ............................................ 17
  1.4 Virulence factors in C. concisus .................................................................................... 18
     1.4.1 Motility ................................................................................................................ 19
     1.4.2 Toxins .................................................................................................................. 20
     1.4.3 Adhesion and invasion ......................................................................................... 25
  1.5 Antibiotic Resistance ................................................................................................. 29
  1.6 Biofilm formation by Campylobacter spp. .................................................................... 30
     1.6.1 Mechanism of biofilm formation ........................................................................ 33
     1.6.2 Cell to cell signalling (Quorum-sensing) ............................................................ 35
  1.7 Role of luxS in QS ....................................................................................................... 39
     1.7.1 The structure and activity of LuxS ....................................................................... 39
<table>
<thead>
<tr>
<th>1.7.2</th>
<th>Metabolic role of LuxS</th>
<th>42</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.7.3</td>
<td>LuxS and growth</td>
<td>44</td>
</tr>
<tr>
<td>1.7.4</td>
<td>LuxS and biofilm formation</td>
<td>45</td>
</tr>
<tr>
<td>1.7.5</td>
<td>LuxS and chemotaxis</td>
<td>45</td>
</tr>
<tr>
<td>1.7.6</td>
<td>LuxS and motility</td>
<td>46</td>
</tr>
<tr>
<td>1.7.7</td>
<td>LuxS and cellular adherence and invasion</td>
<td>48</td>
</tr>
<tr>
<td>1.7.8</td>
<td>LuxS and virulence</td>
<td>49</td>
</tr>
<tr>
<td>1.7.9</td>
<td>LuxS and haemolysis</td>
<td>50</td>
</tr>
</tbody>
</table>

1.8 Molecular typing of *C. concisus* .......................................................... 50

1.8.1 Genetic diversity of *C. concisus* ......................................................... 50

1.8.2 Typing of *C. concisus* on the basis of phenotypic characteristics .......... 53

1.9 Genome analysis of *Campylobacter* spp. .................................................... 54

1.9.1 Whole genome sequence of *C. concisus* .................................................. 55

1.9.2 Comparative genomic analyses of *C. concisus* and *C. jejuni* ................. 56

1.9.3 The ‘pan and core’ genome of *C. concisus* ............................................. 61

1.9.4 Plasmids in *C. concisus* ............................................................................ 62

1.9.5 Phylogenetic analysis based on ribosomal RNA (rRNA) genes .................. 63

1.10 Aims of the study............................................................................................. 63

Chapter Two: General Materials and Methods .................................................... 64

2.1 General Procedures .......................................................................................... 64

2.2 Bacteriological Methods................................................................................... 64

2.2.1 General materials ....................................................................................... 64

2.2.2 General equipment ....................................................................................... 68

2.2.3 General solutions ......................................................................................... 69

2.2.4 Antibiotic stock solutions ............................................................................. 71

2.2.5 Bacteriological media .................................................................................. 71

2.2.6 Bacterial isolates and plasmids ................................................................... 74

2.2.7 Bacterial growth condition .......................................................................... 74

2.2.8 Storing bacterial strains ................................................................................. 75

2.2.9 Measurement of bacterial cell concentrations ................................................. 75

2.3 Microscopy ........................................................................................................ 75

2.3.1 Phase contrast microscopy ............................................................................ 75

2.3.2 Confocal laser scanning microscopy (CLSM) ............................................... 76
2.3.3 Scanning electron microscopy (SEM) ................................................................. 76
2.4 DNA Techniques ................................................................................................. 77
  2.4.1 Genomic DNA extraction and purification ..................................................... 77
  2.4.2 Plasmid DNA extraction and purification ....................................................... 78
  2.4.3 Rapid boiled DNA extraction method ............................................................. 78
  2.4.4 Quantification of DNA concentration ........................................................... 78
  2.4.5 Polymerase Chain Reaction (PCR) ................................................................. 79
  2.4.6 DNA manipulation .......................................................................................... 81
  2.4.7 Preparation of competent \textit{E. coli} DH5\textalpha\ cells ...................................... 82
  2.4.8 Preparation of \textit{C. concisus} competent cells ............................................. 84
  2.4.9 Genetic complementation ............................................................................. 86
2.5 Tissue Culture Methods ...................................................................................... 86
  2.5.1 Tissue culture media and solutions ............................................................... 87
  2.5.2 Tissue culture techniques ............................................................................. 88
2.6 Whole Genome Sequencing (WGS) of \textit{C. concisus} strains .............................. 90
  2.6.1 Preparation of the genomic DNA from \textit{C. concisus} .................................. 90
  2.6.2 WGS in Ion Proton Platform ................................................................. 91
  2.6.3 WGS in Illumina Platform ................................................................. 95
  2.6.4 Sequencing data analysis and quality control ............................................. 98
  2.6.5 \textit{De novo} genome assembly ...................................................................... 98
  2.6.6 Annotation and gene prediction .................................................................. 98
  2.6.7 Plasmid identification and bioinformatics analysis ..................................... 98
Chapter Three : Investigation of \textit{Campylobacter concisus} for biofilm properties .... 99
  3.1 Introduction ...................................................................................................... 99
  3.2 Materials and Methods .................................................................................. 101
    3.2.1 Bacterial strains ....................................................................................... 101
    3.2.2 Molecular identification ........................................................................... 102
    3.2.3 Screening for biofilm-forming \textit{C. concisus} by quantitative crystal violet
        (CV) assay .................................................................................................... 104
    3.2.4 Phenotypic characterization of biofilm produced by \textit{C. concisus} ...... 105
    3.2.5 Analysis of \textit{luxS} in \textit{C. concisus} genome ............................................. 107
    3.2.6 Molecular detection of \textit{luxS} from \textit{C. concisus} .................................. 107
  3.3 Results ............................................................................................................ 108
    3.3.1 Molecular identification ........................................................................... 108
3.3.2 Screening of biofilm-forming *C. concisus* by Quantitative Crystal Violet (CV) assay ................................................................. 109
3.3.3 Phenotypic characterization of *C. concisus* biofilm ................. 111
3.3.4 Phylogeny of luxS in *C. concisus* ........................................... 118
3.3.5 Molecular detection of luxS from *C. concisus* ......................... 121
3.4 Discussion .................................................................................. 125

Chapter Four: Knock-Out mutagenesis of the luxS gene in *Campylobacter concisus* and characterization of ΔluxS mutants ................................................................. 131

4.1 Introduction ................................................................................. 131
4.2 Materials and Methods ................................................................. 132
  4.2.1 Primer design ........................................................................ 132
  4.2.2 Bacteria and plasmids ............................................................. 133
  4.2.3 Genomic DNA extraction ....................................................... 134
  4.2.4 PCR optimization .................................................................. 134
  4.2.5 Cloning of PCR product into primary vector (pCR 2.1 TOPO) .... 136
  4.2.4 Cloning of luxS into pBluescript vector .................................. 137
  4.2.7 Creation of *C. concisus* ΔluxS mutant strain ......................... 139
4.3 Transformation into *C. concisus* .................................................. 140
  4.3.1 Natural transformation of *C. concisus* .................................. 140
  4.3.2 Electroporation of *C. concisus* ............................................. 141
4.4 Genetic complementation of ΔluxS mutant strain ......................... 142
  4.4.1 Construction of a shuttle vector .............................................. 142
  4.4.2 Transformation of shuttle vector p3316 BC luxS .................... 144
4.5 Phenotypic characterization of ΔluxS-O17 ..................................... 144
  4.5.1 Quantitative crystal violet assay ........................................... 144
  4.5.2 Motility assay ........................................................................ 144
  4.5.3 Adhesion and Invasion assay ............................................... 145
  4.5.4 Phenotypic characterization of biofilm by microscopy .......... 146
4.6 Results ....................................................................................... 146
  4.6.1 Creation of *C. concisus* ΔluxS mutant .................................. 146
  4.6.2 Genetic complementation ....................................................... 154
  4.6.3 Phenotypic characterization of the ΔluxS mutant .................... 154
4.7 Discussion .................................................................................. 165
Chapter Five: Comparative Genomics of selected oral and intestinal Campylobacter concisus strains

5.1 Introduction .................................................................................................................. 170
5.2 Materials and Methods ............................................................................................... 170
  5.2.1 C. concisus strain selection for WGS ................................................................. 172
  5.2.2 DNA extraction for WGS .................................................................................... 172
  5.2.3 Whole genome sequencing (WGS) ................................................................. 172
  5.2.4 Draft genome assembly ...................................................................................... 173
  5.2.5 Pairwise comparison of C. concisus strains and analysis ............................... 175
  5.2.6 Annotation of the assembled C. concisus strains and analysis ...................... 175
  5.2.7 Plasmid detection, mapping and construction ..................................................... 176
  5.2.8 Pan- and core-genome analysis ....................................................................... 176
  5.2.9 Enrichment analysis using blast2GO ............................................................... 177
  5.2.10 Detection of clusters of orthologous gene (COG) in four C. concisus strains ............................................................................................................................................................................. 179
  5.2.11 Genes/proteins associated with virulence ....................................................... 179
5.3 Results .......................................................................................................................... 180
  5.3.1 Confirmation of C. concisus DNA quality for genome sequencing ............... 180
  5.3.2 Whole genome sequencing ............................................................................... 182
  5.3.3 Genome assembly of C. concisus strains .......................................................... 185
  5.3.4 Annotation of the assembled Campylobacter concisus strains and analysis ............................................................................................................................................................................. 189
  5.3.5 Investigation of C. concisus 13826 plasmid genes in the studied C. concisus strains ............................................................................................................................................................................. 192
  5.3.6 Plasmid detection, mapping and construction .................................................. 195
  5.3.7 Pan- and core-genome analysis ....................................................................... 198
  5.3.8 Enrichment analysis using blast2GO ............................................................... 209
  5.3.9 Detection of orthologous clusters in four C. concisus genome ....................... 211
  5.3.10 Genes/proteins associated with virulence ....................................................... 214
5.4 Discussion ..................................................................................................................... 217

Chapter Six: Analysis of the ribosomal RNA (rrn) operon of Campylobacter concisus

6.1 Introduction ................................................................................................................ 224
6.2 Materials and methods .............................................................................................. 224
  6.2.1 Source of rrn sequences ................................................................................... 224
6.2.2 In silico analysis of *rrn* operons in *C. concisus* ........................................226
6.2.3 Phylogenetic analysis based on ribosomal RNA genes .................................226
6.3 Results .................................................................................................................227
  6.3.1 Intraspecies sequence variability of rRNA genes .........................................227
  6.3.2 Indels in *rrn* operon of *C. concisus* ...............................................................227
  6.3.3 Indels present in the post-5S region .................................................................229
  6.3.4 Indels present in 23S rRNA genes .................................................................229
  6.3.5 Indels present in ITS region in rRNA genes ....................................................230
  6.3.6 Indels present in 16S rRNA genes .................................................................235
  6.3.7 Phylogenetic analysis of *rrn* operon of *C. concisus* ....................................235
6.4 Discussion ............................................................................................................240

Chapter Seven: General Discussion ...........................................................................245
References ..................................................................................................................256
Appendices ................................................................................................................276
List of Figures

Figure 1.1: Phylogenetic reconstruction of the family *Campylobacteraceae* based on 16S rRNA sequence.................................................................2

Figure 1.2: Notification rates of campylobacteriosis in Australia, 2011, by age group and sex (OzFoodNet, 2011). .................................................................4

Figure 1.3: The role of Proteobacteria in the postulated pathogenesis of IBD.............13

Figure 1.4: Proposed mechanisms of pathogenesis used by emerging *Campylobacter* spp. to colonize the intestinal tract or to spread to systemic sites.................21

Figure 1.5: Host attachment and invasion by *Campylobacter concisus* UNSWCD.......27

Figure 1.6: Phases of *in-vivo* biofilm development...........................................34

Figure 1.7: Summary of genetic components important in biofilm formation.............36

Figure 1.8: AI-2 signalling molecules and receptors..............................................38

Figure 1.9: Crystal structure of *Haemophilus influenzae* LuxS.............................41

Figure 1.10: The AMC (activated methyl cycle) and synthesis of autoinducers...........43

Figure 1.11: Circular layout of seven assembled genomes against the *C. concisus* 13826........................................................................................................58

Figure 1.12: Dendogram of published *C. concisus* genomes based on genomic blast....59

Figure 1.13: The core genomes of *C. concisus* and *C. jejuni* ............................59

Figure 2.1: Overview of the genomic DNA library preparation using the Ion Xpress™ Plus gDNA Fragment Library Preparation kit...........................................92

Figure 2.2: Illumina gDNA fragment library preparation overview..........................96

Figure 3.1: PCR amplification of genomic DNA from *C. concisus* strains using primer pair 2 (CON1 and MUC1) for 23s rRNA gene sequence of *C. concisus* 13826........................................................................................................110

Figure 3.2: Crystal violet quantitative assay for screening of biofilm formation by *C. concisus* strains isolated from patients with gastritis ..........................112

Figure 3.3: Crystal violet quantitative assay for screening of biofilm formation by *C. concisus* strains isolated from oral cavity of healthy persons.........................113

Figure 3.4: Biofilm development by *C. concisus* RMIT-O17 examined by phase contrast microscopy.................................................................115

Figure 3.5: Biofilms of *C. concisus* RMIT-O17 on glass cover slip observed by CLSM at different time of incubation.......................................................117

Figure 3.6: SEM images of *C. concisus* mature biofilms........................................119
Figure 3.7: Biofilm complex produced by RMIT-O17 observed at 20,000x magnification

Figure 3.8: Nucleotide sequence alignment analysis to design PCR primer pairs for amplification of the luxS gene from C. concisus

Figure 3.9: Phylogenetic analysis of luxS genes for the 13 Campylobacter concisus strains showing two clades comprised of genomospecies A and genomospecies B

Figure 3.10: PCR amplification of genomic DNA from C. concisus strains using primer pair 4 for the luxS gene

Figure 4.1: Predicted genetic map of the shuttle vector p3316 BC luxS for complementation

Figure 4.2: Alignment of RMIT-O17 LuxS with C. concisus 13826 LuxS protein using the CLUSTAL O algorithm (1.2.2) (www. ebi.ac.uk)

Figure 4.3: Flow chart showing the strategy adopted in creating knock-out of luxS in RMIT-O17

Figure 4.4: Amplification of flanking regions of luxS, a 1,380 bp PCR product from C. concisus RMIT-O17

Figure 4.5: Digestion of pCR 2.1-TOPO/ luxS with EcoRI reveals fragments at the expected sizes of 3.9 kb and 1,380 bp

Figure 4.6: Digestion of pBK luxS by restriction enzyme KpnI and SacI produced fragments at the expected sizes of 2.9 kb and 1,380 bp

Figure 4.7: Inverse PCR of pBK luxS produced a 4.3 kb PCR product

Figure 4.8: Confirmation of kanamycin insert in pBK luxS kan by restriction enzyme digestion with BamHI/ KpnI and SacI

Figure 4.9: PCR showing the predicted size of luxS knockout in pBK luxS kan and the size of wild type luxS

Figure 4.10: Optimized PCR amplification of flanking regions of luxS, a 1380 bp PCR product from RMIT-O17

Figure 4.11: Digestion of pBC luxS by restriction enzyme XbaI and SacI produced fragments at the expected sizes of 3.4 kb and 1,380 bp

Figure 4.12: Inverse PCR for p3316 in RCH 26 produced a 3.3 kb PCR product

Figure 4.13: Reduced biofilm formation by C. concisus luxS mutant

Figure 4.14: Reduced motility by C. concisus luxS mutant. RMIT-O17 and ΔluxS-O17 were subjected to motility assay on semi-solid agar (0.4% BBA)

Figure 4.15: Adherence of RMIT-O17 and ΔluxS-O17 to INT 407 epithelial cells showing no statistically significant difference
Figure 4.16: Invasion of RMIT-O17 and ΔluxS-O17 to INT 407 epithelial cells showing the mutant exhibited significantly less invasion that the wild type. 

Figure 4.17: Invasive index of RMIT-O17 and ΔluxS-O17 to INT 407 epithelial cells.

Figure 4.18: Biofilm of RMIT-O17 (A) and ΔluxS-O17 (B) by CLSM at 200x magnification and 3.4 zooming showing less biofilm production by luxS-mutant compared to the wild type.

Figure 4.19: Biofilm of RMIT-O17 (A) and ΔluxS-O17 (B) by SEM at 5000x magnification showing less biofilm production by luxS-mutant compared to the wild type.

Figure 5.1: Gel electrophoresis of gDNA extracted from the four C. concisus strains used for whole genome sequencing showing the presence of plasmids only in RCH 26.

Figure 5.2: Electropherogram summary of (A) RMIT-JF1; (B) RCH 26 genomic DNA libraries.

Figure 5.3: Pairwise comparisons of RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17 with C. concisus 13826 using the Mauve alignment tool.

Figure 5.4: Pairwise comparisons of RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17 with C. concisus ATCC 33237T using the Mauve alignment tool.

Figure 5.5: Subsystem statistics and feature counts of AUS22-Bd2 (A), RMIT-JF1 (B), RMIT-O17 (C) and RCH 26 (D) genomes obtained from the Seed Viewer.

Figure 5.6: A sequence based comparison of RMIT JF1, AUS22-Bd2, O17 and RCH 26 annotated genomes with C. concisus ATCC 33237T (A) and C. concisus 13826 (B).

Figure 5.7: Inverse PCR performed on RCH 26 gDNA to confirm plasmids in contig 17 and contig 20 showing the size of PCR products by gel electrophoresis.

Figure 5.8: A circular representation of plasmid sequences present in RCH26 genome. A: pCCON20 (3,316 bp), B: pCCON22 (22,523 bp).

Figure 5.9: Core and pan genome plot of six C. concisus strains using clustering tools-USEARCH.

Figure 5.10: COG distribution of core, accessory and unique genes in C. concisus 13826, ATCC 33237, RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17.

Figure 5.11: Detail COG distribution of core, accessory and unique genes in C. concisus 13826, ATCC 33237, RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17.

Figure 5.12: KEGG distribution of core, accessory and unique genes in C. concisus 13826, ATCC 33237, RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17.
Figure 5.13: Detail KEGG distribution of core, accessory and unique genes in *C. concisus* 13826, ATCC 33237, RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17 .......................................................... 207

Figure 5.14: Phylogenetic analysis by BPGA 1.2 pipeline using genomes of *C. concisus* 13826, ATCC 33237, RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17 based on: (A) concatenated core genes (B) binary pan-matrix .............................................. 208

Figure 5.15: Enrichment of Gene Ontologies for core and pan genes across *C. concisus* 13826, ATCC 33237, RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17 genomes using Fisher’s Exact Test. ................................................................. 210

Figure 5.16: Venn diagram generated using OrthoVenn web based program showing the distribution of shared gene families (orthologous clusters) among RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17 genomes. ................................................................. 212

Figure 5.17: Multiple sequence alignment of *zot* found in RCH 26 and RMIT-JF1 and presence of valine at aa position 270 is indicated in the box. ......................... 216

Figure 6.1: Map of ribosomal RNA (*rrn*) operons from 12 *Campylobacter concisus* strains showing pre-16S, 16S rRNA (red), ITS, tRNAile (light blue), tRNAala (orange), 23S rRNA (blue), 5S rRNA (purple) genes and post-5S regions (shown on top below the consensus). ........................................ 232

Figure 6.2: Indel subtypes with genomospecies A and B from 15 sequences shown in Figure 6.1. ................................................................. 233

Figure 6.3: Evolutionary relationships of 15 sequences of *rrn* operon from 12 *C. concisus* strains .................................................................................. 237

Figure 6.4: Evolutionary relationship among 15 sequences of 16S rRNA gene from 12 *C. concisus* strains ................................................................. 238

Figure 6.5: Evolutionary relationship among 15 sequences of 23S rRNA gene from 12 *C. concisus* strains ................................................................. 239
List of Tables

Table 1.1: Distribution of *Campylobacter* and related species isolated from the stool of children with diarrhea at the Red Cross Children’s Hospital, Cape Town, South Africa, from 1 October, 1990 to 30 June, 2007. ..........................10

Table 1.2: *Campylobacter concisus* genome sequences available in NCBI database....57

Table 3.1: Primer sequences used in this chapter..........................................................103

Table 4.1: The primers used in this study are listed (5’-3’). ........................................135

Table 4.2: luxS and the flanking genes of nine *C. concisus* strains available in NCBI database and their genomospecies. .................................................................148

Table 5.1: Primer sequences used for inverse PCR of plasmids in this chapter. ............178

Table 5.2: Quantification of RMIT-JF1 and RCH 26 prepared libraries by High Sensitivity DNA Assay using Agilent 2100 Bioanalyzer. .................................183

Table 5.3: Summary of the sequencing data of RMIT-JF1 and RCH 26 reported from the Ion Proton sequencing platform. ..............................................................183

Table 5.4: Information on whole genome sequences of *C. concisus* strains used in this study. ..................................................................................................................186

Table 5.5: Conservation of genes from *C. concisus* 13826 plasmids across RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17 using orthoVenn......................................194

Table 5.6: Distribution of core, accessory, unique and exclusively absent genes among *C. concisus* 13826, ATCC 33237, RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17 using USEARCH clustering tool ..................................................202

Table 5.7: Statistical analysis of the pan and core genome plot of *C. concisus* 13826, ATCC 33237, RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17..........................203

Table 5.8: Counts of total proteins, clusters and singletons present in AUS22-Bd2, RCH 26, RMIT-JF1 and RMIT-O17 genomes. .........................................................213

Table 5.9: Presence of virulence related genes among RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17 genomes .................................................................215

Table 6.1: *C. concisus* strains used in this study, their genomospecies, *rrn* names, types and accession numbers. .................................................................228
Summary

*Campylobacter concisus* is a fastidious, hydrogen-requiring bacterium normally found in the human oral cavity, and is currently considered as an emerging pathogen. It has been isolated from gingivitis, periodontitis, foot ulcers, gastritis, and from intestinal biopsies of patients with inflammatory bowel disease (IBD). It is a heterogeneous species of phenotypically indistinguishable strains belonging to different genomospecies. *C. concisus* is also known to produce biofilms on glass, stainless steel, and polystyrene plastic. However, there are little data available on its biofilm properties. The regulation of biofilm formation has been linked to the signalling protein “LuxS” in many oral pathogens. In this study *C. concisus* biofilms were phenotypically characterized and the role of the *luxS* gene in biofilm formation was investigated. Comparative genomic analysis of selected oral and intestinal *C. concisus* strains was also performed to improve the knowledge in the present literature.

Biofilm formation by 14 clinical and 19 oral *C. concisus* strains was assessed by the crystal violet assay. All tested *C. concisus* strains were capable of producing biofilms in different levels, with oral strains being the highest producers. The biofilms were phenotypically characterized by phase contrast microscopy, confocal laser scanning microscopy (CLSM), and scanning electron microscopy (SEM). Different morphological stages of biofilm formation (attachment, maturation and dispersion) were observed by phase contrast microscopy. Completely developed biofilms were observed by CLSM on day five of biofilm formation with a mixture of dead and live bacteria within its structure. In addition, the aggregation of the biofilm and the presence of extracellular polymeric substances (EPS) were confirmed by SEM.
A luxS PCR product (309 bp) was amplified from all clinical and oral strains. Furthermore, to assess the role of LuxS in biofilm formation, a luxS mutant was created by inserting a kanamycin cassette within the luxS gene of RMIT-O17, the highest biofilm-forming strain. RMIT-O17 and the luxS-mutant were subjected to different phenotypic tests. A significant reduction ($p < 0.05$) in the terms of biofilm formation, motility, and invasion of the intestinal epithelial INT 407 cells were observed in the luxS-mutant compared to the wild type. However, no significant difference was found between the adherence properties of the luxS-mutant and the parental strain.

As a heterogenic bacterium C. concisus can be typed into at least two distinct genomospecies (A & B). In this study comparative genomic analysis was performed on the genomes of four C. concisus strains, all belonged to genomospecies A. Those are the genomes of RMIT-O17, along with an oral isolate from an individual with Crohn’s disease (RMIT-JF1) and the two intestinal strains, RCH 26, a faecal strain from a child with gastroenteritis and AUS22-Bd2, a unique strain isolated from the duodenum of an IBD patient. In addition, the genomes of two reference strains C. concisus 13826 and ATCC 33237$^T$, available from the National Center for Biotechnology Information (NCBI), were included in subsequent comparative genomic analyses. The sizes of the sequenced four genomes ranged from 1.83-1.94 Mbp (1,859-2,048 encoded proteins) with the genome of RMIT-JF1 being the largest.

Pairwise comparisons of the genomes of RMIT-O17, RMIT-JF1, RCH 26 and AUS22-Bd2 with C. concisus 13826 and ATCC 33237$^T$ were performed using Mauve alignment tools. The comparison with C. concisus 13826 showed a very high level of gene shuffling (new gene combinations for genetic variation), while a high level of similarity and contiguity was observed with ATCC 33237$^T$. Two plasmids, 22 kb and 3.3 kb, were
detected in the genome of RCH 26 yet the other three genomes did not contain any plasmid. The 3.3kb plasmid was a unique, high copy number plasmid with no similarity to other *C. concisus* genome sequences available in the database.

The pan and core genomes of the four sequenced *C. concisus* strains including *C. concisus* 13826 and ATCC 33237^T^ consisted of 2,790 and 1,463 protein coding genes, respectively. The clusters of orthologous groups (COG) of protein distribution patterns showed the involvement of more core genes in amino acid metabolism and transport; energy production and conversion and translation; ribosomal structure and biogenesis. Pan genes appeared to be enriched in replication, recombination and repair, in addition to cell wall/membrane/envelope biogenesis related functions. This was also supported by the Blast2GO enrichment analysis, where the enrichment of RNA processes, and metabolic and biosynthetic processes were observed within the core genome of *C. concisus*. On the other hand, in the pan genome enrichment of defence responses, DNA-related processes such as DNA integration and DNA restriction-modification suggests that several cellular defence mechanisms do exist in these *C. concisus* strains to probably survive phage attacks. In addition other genes which could be related to phenotypic characteristics such as invasiveness, adherence and motility were identified in the four sequenced *C. concisus* genomes.

Phylogenetic trees were generated based on the ribosomal RNA (*rrn*) operon from 12 *C. concisus* strains. The regions from each *rrn* operon produced different phylogenetic trees demonstrating sequence differences between these strains in relation to the 5S rRNA, 16S rRNA, 23S rRNA, and other intergenic regions, which supports the heterogenic nature of this species. Furthermore, *C. concisus* strains were classified into the two known
genomospecies (A & B) based on the indels present within each operon. Within the \textit{rrn} operon, the 23S rRNA gene was suggested as a reliable region for \textit{C. concisus} typing.

From the findings of this study it can be concluded that the \textit{C. concisus} oral strains are higher biofilm producers than the intestinal strains, which could be a survival mechanism in their normal habitat. The LuxS plays a role in biofilm formation and other virulence properties in \textit{C. concisus}; however, this is yet to be confirmed by genetic or chemical complementation. In addition, genomic analysis of the genomes of oral and intestinal \textit{C. concisus} strains indicated that there is a significant difference in gene content in these strains depending on the isolation site and the clinical history of the host, rather than the genomospecies.
Publications

Journal Publications


Conference Proceedings


- **Oral presentation** presented at ‘*Campylobacter, Helicobacter* and related organisms’ CHRO-2015, 1st-5th November 2015, Rotorua, New Zealand, entitled as ‘Whole genome sequences of oral and intestinal *Campylobacter concisus* strains and molecular typing using the ribosomal RNA operon (rrn)’ by **Mohsina Huq**, Hao Van, Volker Gurtler, Khaled Allemailem, Eltaher Elshagmani, Peter Smooker and Taghrid Istivan.


Abbreviations

<table>
<thead>
<tr>
<th>Term</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>alpha</td>
</tr>
<tr>
<td>Ω</td>
<td>ohms</td>
</tr>
<tr>
<td>λ</td>
<td>lambda phage DNA</td>
</tr>
<tr>
<td>µM</td>
<td>micromole</td>
</tr>
<tr>
<td>mM</td>
<td>millimole</td>
</tr>
<tr>
<td>min</td>
<td>minutes</td>
</tr>
<tr>
<td>µm</td>
<td>micrometer</td>
</tr>
<tr>
<td>µg</td>
<td>microgram</td>
</tr>
<tr>
<td>µl</td>
<td>microliter</td>
</tr>
<tr>
<td>µF</td>
<td>micro Faraday</td>
</tr>
<tr>
<td>aa</td>
<td>amino acid</td>
</tr>
<tr>
<td>Amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Ampicillin resistant</td>
</tr>
<tr>
<td>AFLP</td>
<td>amplified fragment length polymorphism analysis</td>
</tr>
<tr>
<td>AGRF</td>
<td>Australian Genome Research Facility</td>
</tr>
<tr>
<td>ATCC</td>
<td>American type culture collection</td>
</tr>
<tr>
<td>BB</td>
<td>Brucella broth</td>
</tr>
<tr>
<td>BHI</td>
<td>brain heart infusion broth</td>
</tr>
<tr>
<td>BLAST</td>
<td>basic local alignment sequence tool</td>
</tr>
<tr>
<td>Bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>CA</td>
<td>Columbia agar base</td>
</tr>
<tr>
<td>CB</td>
<td>Columbia broth</td>
</tr>
<tr>
<td>CD</td>
<td>Crohn’s disease</td>
</tr>
<tr>
<td>CFU</td>
<td>colony forming units</td>
</tr>
<tr>
<td>CLO</td>
<td>Campylobacter like organism</td>
</tr>
<tr>
<td>CO₂</td>
<td>carbon dioxide</td>
</tr>
<tr>
<td>°C</td>
<td>degrees Celsius</td>
</tr>
<tr>
<td>DMEM</td>
<td>dulbecco’s modified Eagle’s medium</td>
</tr>
<tr>
<td>ddH₂O</td>
<td>deionised distilled water.</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNase</td>
<td>deoxyribonuclease</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxynucleotide triphosphate</td>
</tr>
<tr>
<td>EB</td>
<td>Elution buffer</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediamine tetra acetic acid</td>
</tr>
<tr>
<td>EICC</td>
<td>Enteric invasive <em>C. concisus</em></td>
</tr>
<tr>
<td>EtBr</td>
<td>Ethidium Bromide</td>
</tr>
<tr>
<td>FBS</td>
<td>foetal bovine serum</td>
</tr>
<tr>
<td>g</td>
<td>gram</td>
</tr>
<tr>
<td>Gal</td>
<td>Galactose</td>
</tr>
<tr>
<td>HEPES</td>
<td><em>N</em>-2-hydroxyethylpiperazine-<em>N</em>-2-ethane sulphonic acid</td>
</tr>
<tr>
<td>HBA</td>
<td>Columbia agar base with 5% horse blood agar</td>
</tr>
<tr>
<td>HREC</td>
<td>Human Research Ethics Committee</td>
</tr>
<tr>
<td>h</td>
<td>hour</td>
</tr>
<tr>
<td>IBD</td>
<td>inflammatory bowel disease</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl-β-D-thiogalactosidase</td>
</tr>
<tr>
<td>Kb</td>
<td>kilobase pairs</td>
</tr>
<tr>
<td>Km&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Kanamycin resistance</td>
</tr>
<tr>
<td>M</td>
<td>molar</td>
</tr>
<tr>
<td>mA</td>
<td>Milli ampere</td>
</tr>
<tr>
<td>MCS</td>
<td>Multiple cloning site</td>
</tr>
<tr>
<td>mg</td>
<td>milligram</td>
</tr>
<tr>
<td>MLST</td>
<td>multilocus sequence typing</td>
</tr>
<tr>
<td>min</td>
<td>minute</td>
</tr>
<tr>
<td>MOI</td>
<td>the multiplicity of infection</td>
</tr>
<tr>
<td>ng</td>
<td>nanogram</td>
</tr>
<tr>
<td>N&lt;sub&gt;2&lt;/sub&gt;</td>
<td>Nitrogen</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>pH</td>
<td>negative logarithm to the base ten of the concentration of hydrogen ions</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>qPCR</td>
<td>quantitative PCR</td>
</tr>
<tr>
<td>RCH</td>
<td>The Royal Children's Hospital</td>
</tr>
<tr>
<td>RAPD</td>
<td>randomly amplified polymorphic DNA</td>
</tr>
<tr>
<td>rDNA</td>
<td>ribosomal deoxyribonucleic acid</td>
</tr>
<tr>
<td>sec</td>
<td>second</td>
</tr>
<tr>
<td>SD</td>
<td>standard deviation</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SOC</td>
<td>Super optimal broth with catabolite repression</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-acetate-EDTA</td>
</tr>
<tr>
<td>UC</td>
<td>ulcerative colitis</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>V</td>
<td>voltage</td>
</tr>
<tr>
<td>v/v</td>
<td>volume per volume</td>
</tr>
<tr>
<td>WGS</td>
<td>whole-genome sequence</td>
</tr>
<tr>
<td>w/v</td>
<td>weight per volume</td>
</tr>
<tr>
<td>X-gal</td>
<td>5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside</td>
</tr>
</tbody>
</table>
Chapter 1: General Introduction

Chapter One

General Introduction

1.1 The genus *Campylobacter*

The genus *Campylobacter* is a member of the family *Campylobacteraceae*, which consists of the genera *Campylobacter* (30 taxa), *Arcobacter* (17 taxa) and *Sulfurospirillum* (7 taxa) (Figure 1.1) (Lastovica et al., 2014). Phylogenetic trees drawn from alignments of 16S rDNA sequences of *Campylobacter* spp. were all found to contain three distinct species groups (Gorkiewicz et al., 2003). The first group contains *Campylobacter* spp. with proven associations for both human and animal periodontal cavities such as *C. hominis*, *C. gracilis*, *C. sputorum*, *C. curvus*, *C. rectus*, *C. showae* and *C. concisus*. The second group consists of *C. fetus*, *C. hyointestinalis* and *C. mucosalis*, which are associated with disease in farm animals although the first two have also been occasionally implicated in human diseases. The third group consists of *C. jejuni*, *C. coli*, *C. lari*, *C. helveticus*, and *C. upsaliensis*, all known to cause gastroenteritis in humans except *C. helveticus* (Lawson et al., 1999).

*Campylobacter* spp. are mostly motile, slender, spirally curved rods or helical rods often displaying “gull wing” morphology with a size range of 0.5–0.8 µm by 0.5–5.0 µm. They grow in a microaerophilic atmosphere and have optimal O₂ of 5–10 % and CO₂ of 3–5 %. *C. gracilis*, *C. hyointestinalis*, *C. showae*, and *C. sputorum bv. faecalis* grow under anaerobic conditions too. *C. concisus*, *C. curvus*, *C. mucosalis*, *C. rectus*, and *C. gracilis* require H₂ or formate for growth. *Campylobacter* spp. have an optimal growth temperature
Figure 1.1: Phylogenetic reconstruction of the family Campylobacteraceae based on 16S rRNA sequence, adapted from Lastovica et al. (2014).
of 30–42°C. Oxidase activity is present in all species except \textit{C. gracilis}, but there is no lipase or lecithinase activity in most \textit{Campylobacter} spp., and the %GC content of their DNA ranges from 29–47 mol%.

### 1.1.1 History and association with diseases

In 1963 Sebald and Veron used the G+C ratio in genomic DNA and Hugh and Leifson’s test for fermentative metabolism proposed the genus—\textit{Campylobacter} (Lastovica \textit{et al.}, 2014). Interest in \textit{Campylobacter} increased dramatically when Butzler \textit{et al.} (1973) and later Skirrow (1977) showed \textit{Campylobacter} spp. were frequently associated with human diarrhea. By the end of 1970’s, it was considered as an established human pathogen. \textit{Campylobacter} spp. are known to be associated with gastroenteritis, Guillain-Barré syndrome, inflammatory bowel diseases, oesophageal diseases, periodontal diseases, celiac disease, cholecystitis, hemolytic-uremic syndrome, pancreatitis, reactive arthritis and colon cancer (Kaakoush \textit{et al.}, 2015a).

Gastroenteritis, a severe form of diarrhoea, is a major clinical condition resulting from \textit{Campylobacter} infection, occurring worldwide especially in children in both developed and developing countries (Allos, 2001; Blaser, 1997; Mead \textit{et al.}, 1999; OzFoodNet, 2011). Figure 1.2 shows the incidence of campylobacteriosis notified cases in Australia in 2011. \textit{Campylobacter} spp. is normally transmitted by consumption of contaminated foods, specially poultry products, unpasteurized milk, and untreated tap, tank and bore water (Friedman \textit{et al.}, 2004; Olson \textit{et al.}, 2008). It can also be transmitted via animals (Neimann \textit{et al.}, 2003; Friedman \textit{et al.}, 2004). Although campylobacteriosis is a self-limiting infection, it costs $1.747 billion per year in the USA alone for treatment including medical costs, productivity losses and valuation of mortality (Batz \textit{et al.}, 2012).
Figure 1.2: Notification rates of campylobacteriosis in Australia, 2011, by age group and sex (OzFoodNet, 2011).
Due to underreported cases, the actual rates of campylobacteriosis are considered to be 10–100 times higher than those reported historically (Guerry et al., 2012). Although C. jejuni and C. coli are mostly responsible for campylobacteriosis (Skirrow, 1994), other emerging Campylobacter spp. including C. gracilis, C. upsaliensis, C. lari and particularly C. concisus also play a role in intestinal disease (Lastovica, 2006; Man, 2011).

1.2 Campylobacter concisus: the emerging pathogen

C. concisus is a Gram negative fastidious hydrogen-requiring spiral or curved shaped rod with a size of (0.5- 1) × (2 - 6) μm (Garrity et al., 2005). C. concisus is motile, driven by a single polar flagellum. Unlike other Campylobacter spp, C. concisus does not have any known primary animal reservoir and probably the human gastrointestinal tract is its only habitat and infections may spread via the inter-personal route only. However, the DNA of C. concisus has been detected in saliva of domestic pets by the PCR-DGGE method (Petersen et al., 2007), and in diarrheic faecal samples from domestic dogs by quantitative PCR (Chaban et al., 2010). C. concisus was isolated from slaughtered porcine samples but not from live animals (Scanlon et al., 2013).

1.2.1 C. concisus in the human oral cavity

In 1981, C. concisus was first recognized and named as a member of the microflora of the oral cavity by Tanner et al. (1981). The oral cavity includes all part of the mouth (lips, buccal mucosa, teeth, gums, tongue, the floor of the mouth below the tongue, and the bony roof of the mouth). It remains unclear whether C. concisus is an oral pathogen, an opportunistic pathogen of inflamed tissues, or is simply a commensal of the oral cavity.
1.2.1.1 *C. concisus in oral cavity of healthy subjects*

*C. concisus* was detected in the human oral cavity by Macuch and Tanner (2000) with other *Campylobacter* spp. such as *C. showae*, and *C. curvus*. It was suggested that these bacteria colonize the oral cavity more frequently than might be expected for transient species and may represent opportunistic pathogens under certain medical conditions. Interestingly, in this study *C. concisus* was isolated in higher proportions from relatively shallow pockets (1-3 mm in depth) (19.7%) and from healthy sites (no redness or bleeding after probing), compared with deeper pockets (3-11 mm deep) (6.6%). Other studies examined the composition of the subgingival microbiota of children, and 45 microbial species were isolated from subgingival plaque of both permanent and deciduous teeth of systemically healthy children. The detection rate of *C. concisus* in permanent teeth was significantly higher than that of deciduous teeth (*p* < 0.001) (Kamma *et al.*, 2000b; Kamma *et al.*, 2000a). The prevalence of *C. concisus* in the human oral cavity was detected by a PCR targeting the 16S rRNA gene and it was found in 100% of saliva samples (11/11) collected from healthy individuals (Petersen *et al.*, 2007). Similar results were found by Dewhirst *et al.* (2010) where defining of the human oral microbiota was attempted by cultivation and culture-independent molecular methods such as 16S rRNA sequencing. *C. concisus* was proposed as a member of the human oral microbiota. In another study, *C. concisus* was isolated from saliva of healthy controls, where 75% (44/59) were culture positive and 97% (57/59) were PCR positive which indicate that *C. concisus* is commonly present in the human oral cavity (Zhang *et al.*, 2010). These collective data show that humans are the natural host of *C. concisus*, with the human oral cavity being the primary colonization site.
1.2.1.2  \textit{C. concisus in gingivitis and periodontitis}

The association of \textit{C. concisus} with human periodontal diseases is well known since 1981 (Tanner \textit{et al.}, 1981; Tanner \textit{et al.}, 1987; Kamma \textit{et al.}, 2000a). \textit{C. concisus} was reported to be associated with periodontitis where higher numbers of \textit{C. concisus} were found to be attached to teeth than other sites of the oral cavity of patients with periodontitis (Haffajee \textit{et al.}, 1984). The immune response to \textit{C. concisus} in periodontal disease was investigated and higher antibody levels against \textit{C. concisus} were detected in periodontally diseased subjects compared to the healthy controls (Taubman \textit{et al.}, 1992). Later on it was reported to be associated with gingivitis, periodontal sites in addition to healthy sites (Tanner \textit{et al.}, 1998).

Furthermore, \textit{C. concisus} was reported to be found in bleeding sites more than non-bleeding site in periodontitis (Kamma \textit{et al.}, 1994; Kamma \textit{et al.}, 2000a; Kamma \textit{et al.}, 2000b). It was also isolated from enlarged lesions of gingivitis (Kamma \textit{et al.}, 1998). In 1999, the same group found that \textit{C. concisus} was more associated with periodontitis in smoker than non-smoker patients (Kamma \textit{et al.}, 1999). The association of \textit{C. concisus} with periodontitis was also supported by significantly higher isolation rates of \textit{C. concisus} when gingival crevicular fluid (inflammatory exudate and a marker of the periodontal disease) of patients was positive for aspartate aminotransferase (AST) compared to patients negative for AST in gingival crevicular fluid (Kamma \textit{et al.}, 2001). AST is released into the extracellular space during cell damage or in cell death (Schmidt and Schmidt, 1985) and high levels of AST in gingival crevicular fluid may indicate active periodontal disease (Sheth and Verma, 2011). Later, \textit{C. concisus} was included into one of the six successional complexes that are believed to be involved in periodontal diseases (Socransky and Haffajee, 2005).
1.2.2 *C. concisus* in acute gastroenteritis

The correlation between *C. concisus* and gastroenteritis was first reported in 1989 by Vandamme *et al.* (1989). A group of unidentified or misidentified human, non-oral clinical *Campylobacter* strains (EF group, 22 strains) were correctly identified by routine immunotyping. They were characterized by numerical analysis of gel electrophoretic protein profiles, immunotyping, DNA hybridizations and sequencing.

*C. concisus* along with other *Campylobacter* spp. such as, *C. upsaliensis*, *C. hyointestinalis*, and *C. fetus* have been reported to cause gastroenteritis, but remained unidentified due to difficulties in conventional culture techniques and because of the use of antibiotics that may inhibit their growth (Linton *et al.*, 1996; Kulkarni *et al.*, 2002; Samie *et al.*, 2007). Table 1.1 shows the isolation rate of *Campylobacter* spp. species in children with diarrhoea in South Africa from 1999 to 2007.

The isolation procedure of *C. concisus* was improved when H₂ was introduced to the microaerophilic incubation conditions and *C. concisus* was isolated from the stools of 2.4% of children (*n* =3165) and 1.5% of adults (*n* =1265) with diarrhoea (Lauwers *et al.*, 1991). However in this study, 20% of children had a co-infection with other enteric pathogens, and more than 25% of children were harbouring *C. concisus* without any enteritis symptoms, indicating that more investigation was required to confirm the role of *C. concisus* in gastroenteritis. In Sweden *C. concisus* was isolated from children with diarrhoea using the filtration technique and 6% of the total children but no adults were infected with *C. concisus* (Lindblom *et al.*, 1995). *C. concisus* strains isolated in the Royal Children’s Hospital in Melbourne in 1995 and 56% of the total CLOs (*Campylobacter* like organisms) were found to be *C. concisus* with the majority of the
patients being under 2 years of age (Russell, 1995) with the majority of the patients being under 2 years of age. Hence, *C. concisus* was reported to be associated with diarrhoea particularly in infants from 0-35 months of age (Istivan, 2005). However, Van Etterijck *et al.* (1996) carried out an experiment using the filtration technique on children with and without diarrhoea and found no significant difference (9% in control and 13.2% in patients). Another study found that atypical *Campylobacter* spp. (other than *C. jejuni* and *C. coli*) comprised 20% of all *Campylobacter* spp. and 0.7% of them were *C. concisus* in children with gastroenteritis (Musmanno *et al*., 1998). A study in Denmark in the year 2000 identified *Campylobacter* spp. in stool samples from diarrhoeal patients and healthy individuals by culture methods and out of the 1,376 samples, 144 (10.4 %) were positive for *Campylobacter* spp. Of those 74 were *C. jejuni* or *C. coli* and 45 were *C. concisus* (Engberg *et al*., 2000). The isolation rate of *C. concisus* was also increased by 23.55% at the Red Cross Children’s Hospital in Cape Town, South Africa after they started using the filtration technique (Lastovica and le Roux, 2000).

*C. concisus* DNA has been detected in stool samples from patients with gastroenteritis in several studies (Maher *et al*., 2003; Samie *et al*., 2007; Vandenberg *et al*., 2013; Huq *et al*., 2014; Underwood *et al*., 2015). Nielsen *et al.* (2013a) has found high incidence of *C. concisus* in patients with gastroenteritis, almost as high as the common *C. jejuni* or *C. coli* in a population based study in Denmark. In the same year, this research group suggested that a polycarbonate filter is superior to the commonly used cellulose acetate filter for detection of *C. concisus* (Nielsen *et al*., 2013c). A significantly higher prevalence of *C. concisus* was detected with the polycarbonate filter (n = 114) compared to the cellulose acetate filter (n = 79) \((p < 0.0001)\).
Table 1.1: Distribution of *Campylobacter* and related species isolated from the stool of children with diarrhea at the Red Cross Children’s Hospital, Cape Town, South Africa, from 1 October, 1990 to 30 June, 2007, adapted from Lastovica and Allos. (2008).

<table>
<thead>
<tr>
<th>Species or subspecies</th>
<th>No</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. jejuni</em> subsp. <em>Jejuni</em></td>
<td>1956</td>
<td>32.57</td>
</tr>
<tr>
<td><em>C. concisus</em></td>
<td>1503</td>
<td>25.02</td>
</tr>
<tr>
<td><em>C. upsaliensis</em></td>
<td>1414</td>
<td>23.54</td>
</tr>
<tr>
<td><em>C. jejuni</em> subsp. <em>doylei</em></td>
<td>431</td>
<td>7.18</td>
</tr>
<tr>
<td><em>H. fenelliae</em></td>
<td>337</td>
<td>5.61</td>
</tr>
<tr>
<td><em>C. coli</em></td>
<td>181</td>
<td>3.01</td>
</tr>
<tr>
<td><em>C. hyointestinalis</em></td>
<td>57</td>
<td>0.95</td>
</tr>
<tr>
<td><em>H. cinaedi</em></td>
<td>51</td>
<td>0.85</td>
</tr>
<tr>
<td>CLO/HLO</td>
<td>31</td>
<td>0.52</td>
</tr>
<tr>
<td><em>A. butzleri</em></td>
<td>20</td>
<td>0.33</td>
</tr>
<tr>
<td><em>C. fetus</em> subsp. <em>fetus</em></td>
<td>9</td>
<td>0.15</td>
</tr>
<tr>
<td><em>C. curvus</em>, <em>C. citrus</em> and <em>H. rappini</em></td>
<td>9</td>
<td>0.15</td>
</tr>
<tr>
<td><em>C. sputorum</em> biovar <em>sputorum</em> and <em>C. lari</em></td>
<td>7</td>
<td>0.12</td>
</tr>
<tr>
<td>Total</td>
<td>6006</td>
<td>100.00</td>
</tr>
</tbody>
</table>

CLO/HLO, *Campylobacter* or *Helicobacter* organisms that could not be fully characterized.
1.2.3 *C. concisus* in chronic gastrointestinal diseases

Inflammatory bowel diseases (IBD) are chronic inflammatory conditions of the gastrointestinal tract (GIT). This condition has been associated with *C. concisus* since 2009 (Zhang *et al.*, 2009). The two major types of IBD are Crohn’s disease (CD) and ulcerative colitis (UC). CD inflammation has a patchy distribution in all layers of the large intestine and additionally in the rectum, whilst UC is characterized by continuous submucosal inflammation occurring only in the colon and rectum. In CD, perianal fistulas and ulcers are common with the presence of granulomas and giant cells while these are rare in UC. Cancer of the colon is a risk in patients with both UC and CD but more common with UC (Ament, 1975).

The role of gastrointestinal microbiota has been investigated for the cause of IBD, specifically the large and diverse phylum ‘Proteobacteria’ of which *C. concisus* is a member. The probable mechanism of IBD is demonstrated in Figure 1.3 (Mukhopadhya *et al.*, 2012).

*C. concisus* was first isolated from stools samples of IBD patients (n = 11) for lectin typing (Aabenhus *et al.*, 2002a), however, the isolated strains from IBD patients could not be classified into a single type. Zhang *et al.* (2009) compared the prevalence of *C. concisus* in patients with CD and controls and found a significantly higher prevalence in children with CD than in controls (*p* < 0.001). This finding was supported in 2010 where *C. concisus* DNA was detected from faecal samples (65%) of CD patients, which was significantly higher than that of in healthy and non-IBD controls (33%) (Man *et al.*, 2010c). The first report of an association of *C. concisus* with UC came in 2011 (Mukhopadhya *et al.*, 2011). In biopsy specimens the prevalence of *C. concisus* DNA was
Figure 1.3: The role of Proteobacteria in the postulated pathogenesis of IBD. (a) Normal intestinal mucosa with resident commensal bacteria including Proteobacteria in the intact mucus layer. The epithelial cells, with intervening tight junctions, in conjunction with the local immune system prevent translocation of bacterial pathogens. (b) Changes that occur as a result of acute enteritis caused by pathogenic Proteobacteria, which could be the primary trigger-event in the pathogenesis of IBD. Acute infection with Campylobacter and Salmonella species usually results in a complete remission of symptoms with return of the luminal microenvironment to normal. (c) A shift in the ratio of pathogenic bacteria (including Proteobacteria) and other commensal bacteria in genetically susceptible individuals. In this setting, adherent-invasive Proteobacteria colonize the lumen, adhere to the epithelial surface and subsequently invade into the lamina propria as a result of defective host recognition and impaired autophagy. (d) Disruptions that occur in IBD. Pathogenic Proteobacteria gain further access through M cells and denuded epithelial surfaces. They replicate unchecked in the phagolysosome of macrophages and initiate proinflammatory cytokines that drive characteristic chronic inflammation in patients with IBD. The proinflammatory cytokines modify the luminal environment with resultant change in the resident bacterial niche and overgrowth of ‘inflammophilic’ pathogenic bacteria that further aggravate the proinflammatory response. Proteobacteria are considered the ‘agent provocateur’ in this inflammatory cascade. Abbreviation: PRR, pattern-recognition receptor, adapted from Mukhopadhyya et al. (2012).
significantly increased \((p = 0.0019)\) in adult UC patients \((33.3\%)\) as compared with controls \((10.8\%)\). Interestingly, higher prevalence of \textit{C. concisus} and \textit{C. ureolyticus} found in biopsy samples from adults with UC suggested that co-infection of these two species of bacteria might be a characteristic in the chronic inflammation in UC. Later in the same year, this finding was supported by another study, where detection of both \textit{Campylobacter} spp. by PCR and isolation of \textit{C. concisus} in patients with UC and CD were significantly higher \((p < 0.05)\) than that in the controls \((\text{Mahendran et al., 2011})\). \textit{C. concisus} were more often detected in descending colonic and rectal biopsies from patients with IBD in comparison to the controls. In another study, the sera of \textit{C. concisus}-positive children with CD were tested for \textit{C. concisus} immunoreactive proteins and flagellin B, ATP synthase F1 alpha subunit and outer membrane protein 18 which were found as common antigens recognized by all CD patients \((\text{Kovach et al., 2011})\).

Ismail \textit{et al.} (2012) used a total of 71 enteric and oral \textit{C. concisus} from eight patients with IBD (four UC and four CD) and six controls for comparison by multilocus sequence typing (MLST), invasion assay, protein analysis and scanning electron microscopy. Interestingly, MLST results showed the majority \((87.5\%)\) of the individuals having IBD were in one cluster compared to the control groups \((28.6\%)\) \((p < 0.05)\). This study provided the first evidence that patients with IBD (both CD and UC) are colonized with specific oral \textit{C. concisus} strains and these strains may undergo natural recombination. In the same year in a retrospective survey in North Denmark, 174 individuals were found positive for \textit{C. concisus} and of those 7 had UC, 5 had CD and 2 had macroscopic colitis at the time of stool collection \((\text{Nielsen et al., 2012})\). Recently it was found by real time PCR that the level of \textit{C. concisus} and exotoxin 9 (a putative virulence factor which may associate with increased survival in the cell) were significantly higher in CD patients.
compared to the healthy controls in faecal samples (Kaakoush et al., 2014). In another study, it was suggested that *C. concisus* strains harbouring the zonula occludens toxin (Zot) may weaken the primary barrier function of the intestine and may trigger the onset and relapse of IBD (Zhang et al., 2014).

### 1.2.4 *C. concisus* in immunocompromised patients

The presence of *C. concisus* in stool samples of immunocompromised patients was found to be in 110 out of 224 *Campylobacter* positive patients, and 26 out of 110 had co-infection with other established pathogens like *Clostridium difficile*, *Giardia*, *Salmonella*, *Entamoeba coli*, *Shigella* and *Yersinia*. Patients with only *C. concisus* (73%) had diarrhoea, bloody stool and fever. This study suggested that *C. concisus* is the causative agent of diarrhoea in immunocompromised patients (Aabenhus et al., 2002b). In a study in the Venda region, South Africa, 10 (3.2%) samples out of 322 were positive for *C. concisus*. A total of six patients had infection with *C. concisus* and among those only two were immunocompromised patients. These results demonstrate the relationship of *C. concisus* with gastroenteritis (Samie et al., 2007). In another study, *C. concisus* has been detected by PCR amplification of bacterial 16S ribosomal RNA (rRNA) from gastric fluid in significantly greater amounts from immunocompromised patients (von Rosenvinge et al., 2013).

### 1.2.5 *C. concisus* in other parts of the human body

Although the majority of research on the correlation of *C. concisus* with human disease has been associated with periodontitis, gastroenteritis, and IBD, there are several studies concluding that *C. concisus* is a potential cause of other human diseases. *C. concisus* was found with many other bacteria on the tongue dorsum and was related to oral
malodour comprising 12.5% of the total anaerobes found on the tongue (Tyrrell et al., 2003). It may also be responsible for osteomyelitis of the sacrum in a patient with diabetes and a sacral decubitus ulcer (Engberg et al., 2005; Johnson and Finegold, 1987). In another study, *C. concisus* and *C. rectus* were isolated from patients with Barrett’s esophagitis (BE), a condition in which gastroduodenal reflux leads to squamous epithelial cells lining the esophagus being replaced with columnar epithelial cells (Macfarlane et al., 2007). As a result patients are at greatly increased risk of esophageal dysplasia and adenocarcinoma. In this study, high level of *C. concisus* and *C. rectus* were found in four (57%) of seven patients and none of the control (0/7) group from endoscopic biopsy samples. *C. concisus* was isolated from both the aspirate and mucosal samples, and was the most prevalent bacterium. The authors hypothesized that pathogenic and putative toxin producing *Campylobacter* spp. could be involved in the initiation, maintenance, or exacerbation of this disease leading to adenocarcinoma (Macfarlane et al., 2007). However, the sample size was not big enough. In a report on the isolation of *Campylobacter* spp. from extra-oro-intestinal abscesses, *C. concisus* was isolated in a polymicrobiota from a brain abscess of a patient who had a history of maxillary sinus carcinoma (present with *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Stenotrophomonas maltophilia*, *Prevotella* species, and *Peptostreptococcus anaerobius*). The patient later died from an intracerebral infection due to chronic sinusitis indicating that *C. concisus* was an opportunistic pathogen in this case (de Vries et al., 2008). *C. concisus* was also isolated from blood samples of HIV patients and children (Lastovica, 2009). A survey of paediatric records indicated the isolation of *C. concisus* from stool specimens of nine liver transplant patients, six patients with biliary atresia and three with renal transplants or renal failure (Lastovica, 2009).
1.3 Identification of C. concisus

1.3.1 Identification of C. concisus by cultural and biochemical properties

*C. concisus* is a hydrogen-requiring, slow-growing fastidious bacterium. It is biochemically inert or inactive. It has been under-reported due to difficulties in isolation and improper identification. Sensitivity to cephalothin and nalidixic acid, growth temperature and colony colour have been used to identify *C. concisus* (On, 1994). Arylsulfatase activity test was used to differentiate *C. concisus* from *C. mucosalis* and *C. upsaliensis* (Lastovica et al., 1994). It was misidentified as *C. mucosalis* when initially isolated from samples other than oral cavity (Vandamme et al., 1989). Now, *C. concisus* is reported more often from patients with diarrhoea and other sites because of improvement of the culture system and use of the stool filtration cellulose acetate filter technique (Bolton et al., 1988). In Cape Town the identification rate increased by 31% because of the use of this technique with hydrogen-enriched environment (Matsheka et al., 2001; Engberg et al., 2000). However, as mentioned earlier, recently Nielsen et al. (2013c) demonstrated the polycarbonate filter is superior to the cellulose acetate filter for detection of *C. concisus*.

1.3.2 Identification of C. concisus by molecular methods

Historically, determining the presence of *C. concisus* relied on conventional methods such as culturing. This technique poses many challenges and can result in false negatives due to several external factors. Molecular biology allows more freedom as well as a higher sensitivity when determining presence and absence of the pathogens. A variety of molecular methods have been approached to detect *C. concisus* since it has been identified. Vandamme et al. (1989) used electrophoretic protein profiles, immunotyping and DNA: DNA hybridization to identify 22 strains named as EF (E Falsen) group 22.
which were identified as *C. concisus*. The 22 strains showed a considerable heterogeneity (42%) with the *C. concisus* type strain. The researchers suggested that SDS-PAGE and immunotyping was an excellent tool for identification of *C. concisus*. In 1995 a *C. concisus* specific PCR was developed from the 23S rDNA region (Bastyns *et al.*, 1995). However, there were some reports that the primers designed for this PCR constantly cross reacted with *C. showae* and *Wolinella succinogenes* and produced a similar size PCR product (Engberg *et al.*, 2000). Again, Marshall *et al.* (1999) developed a two-step identification scheme for *Campylobacter, Arcobacter* and *Helicobacter* based on analysis of the 16S rRNA gene by PCR-RFLP (PCR-restriction fragment length polymorphism) and successfully identified *C. concisus*. In 2001, a PCR assay was developed from a 1.6 kb DNA fragment isolated from *C. concisus* genomic library for molecular identification, where a single PCR product was obtained without any cross reaction from other *Campylobacter* spp. (Matsheka *et al.*, 2001). In 2007 a PCR-DGGE (PCR-denaturing gradient gel electrophoresis) method was developed for detecting *Campylobacter, Helicobacter* and *Arcobacter* from clinical samples and *C. concisus* was identified by this new method (Petersen *et al.*, 2007). Furthermore a nested PCR was developed using *Campylobacter* genus specific primers and in the second step targeting the 16S rRNA gene of *C. concisus* (Man *et al.*, 2010c). This PCR was applied to detect *C. concisus* from children’s stool samples with CD, non-IBD patients and healthy controls. Later, this nested PCR method was used for the detection of *C. concisus* from saliva samples of IBD patients and healthy controls (Zhang *et al.*, 2010).

### 1.4 Virulence factors in *C. concisus*

The currently understood mechanisms for disease-causing *Campylobacter* spp. are colonization, attachment, invasion, and toxin production (Lastovica *et al.*, 2014). After
Chapter 1: General Introduction

entering the host, *Campylobacter* spp. first evade the innate immune system by adhering to the host cell surfaces, and preventing removal by peristaltic movement of the GIT. They then multiply, colonise, and resist host defence mechanisms by expressing a range of virulence factors. Virulence factors studied in *Campylobacter* spp. include the flagella, adhesins, cellular invasion, toxins, phospholipase A, iron-scavenging systems and resistance to oxidation damage (Ketley, 1997). Flagellum is the most studied virulence factor in *C. jejuni*, which is needed for adhesion and colonisation. After adhesion and colonisation to intestinal cell surfaces, *Campylobacter* spp. invade or produce various toxins to damage the epithelial cell lining to interfere with the normal absorptive capacity of the intestine (Ketley, 1997). Until recently, the pathogenic potential of *C. concisus* has remained relatively unexplored as it was mostly related to periodontitis. However, it has now been associated to IBD and gastroenteritis. A significant number of studies have examined potential virulence factors in *C. concisus*. Results from these studies suggest that multiple virulence factors are involved in the pathogenicity of *C. concisus*.

1.4.1 Motility

Flagellum is the motility organelle of bacteria and plays an important role in virulence mechanism as it helps the cell to swim through the highly viscous intestinal mucus and also in attachment to the intestinal epithelial cells (Ramos *et al.*, 2004). The first study to investigate *C. concisus* motility and chemotaxis was done by Paster and Gibbons (1986) with a *C. concisus* oral isolate. The chemotactic response was tested to a wide range of compounds and showed that *C. concisus* 288 exhibited chemotactic response to formate, but it was not chemoattracted to 20 tested sugars, inorganic salts, amino acids and their derivatives, purines and pyrimidines, fatty acids, or natural mixtures such as saliva, serum, mucin and crevicular fluid. The presence of a single polar flagellum
in *C. concisus* was confirmed by scanning electron microscopy (Paster and Gibbons, 1986; Man *et al.*, 2010a). It was shown that with the help of this flagellum, *C. concisus* is able to swim through protective host barriers such as saliva or the intestinal mucus layer and colonise the epithelial layers (Man *et al.*, 2010a). Kaakoush *et al.* (2011b) demonstrated the aggregation and possible positive-mucin chemotaxis of clinical *C. concisus* (human intestinal biopsies), which contrasted the findings of Paster and Gibbons (1986). The possible explanation is the source of isolation was different in two experiments and the previous experiment (Paster and Gibbons, 1986) was performed on one strain only. Kaakoush *et al.* (2011a) also have identified significant variations within the flagellin glycosylation pathways of different *C. concisus* strains, which may affect the pathogenicity. In 2011, *C. concisus* proteins functionally classified as involved in chemotaxis, signal transduction, flagellar motility, surface binding and membrane protein assembly were identified by mass spectrometry (Kovach *et al.*, 2011). Later, Lavrencic *et al.* (2012) have shown that the motility of *C. concisus* increased in media resembling the viscosity of the outer mucus layer of the colon as compared with that in media of lower viscosity.

### 1.4.2 Toxins

*C. concisus* strains have been shown to produce several toxins including cytolytic effect on Chinese hamster ovary (CHO) cells, haemolysins, phospholipase A2 activity, pore forming toxins, and zonula occludens toxin (Zot) (Man *et al.*, 2010a; Istivan *et al.*, 2008; Kalischuk and Inglis, 2011; Zhang *et al.*, 2014).
Figure 1.4: Proposed mechanisms of pathogenesis used by emerging *Campylobacter* spp. to colonize the intestinal tract or to spread to systemic sites. *Campylobacter* are mucosa-associated bacteria that have the capacity to swim through the mucus layer in intestinal epithelium. A number of emerging *Campylobacter* have been shown to attach to and invade intestinal epithelial cells. It is suggested that these bacteria can translocate across the intestinal epithelium by breaking down barrier-associated tight junctions. They might deliver effector proteins into host epithelial cells via a putative T4SS. The S-layer mediates resistance to phagocytosis and serum-killing, possibly by inhibiting stable complement deposition onto bacterial cell surface during the systemic phase of infection. They can also generate a number of toxins, including the tripartite cytolethal distending toxin. Other toxins include cell-bound and secreted hemolysins, which have the capacity to lyse red blood cells. Abbreviations: S-layer, surface layer; T4SS, type IV secretion system, adapted from Man (2011).
1.4.2.1 Cytotoxin

The first study on *C. concisus* toxins was conducted by Musmanno et al. (1998) which reported cytotoxic-like effects on Chinese hamster ovary (CHO) cells and inducing intra-cytoplasmic vacuole formation in INT 407 similar to *H. pylori*. Later these findings were supported by observing stable vacuolating and cytolysis effects on CHO cells in tissue culture (Istivan et al., 2004). Moreover, the ability of *C. concisus* strains to induce cytolethal distending toxin-like effects on monkey kidney (Vero) cells was detected but it was found unrelated to the disease outcome (Engberg et al., 2005).

Further evidence on the cytotoxicity of *C. concisus* isolates was provided by Nielsen et al. (2011), where an elevated level of lactate dehydrogenase release was observed in human HT-29/B6 colon cells infected with *C. concisus*. However, Kalischuk and Inglis (2011) found no significant epithelial cytotoxicity (measured through lactate hydrogenase release) in any *C. concisus* isolates used in the study. The gene of CDT B toxin was also not detected in any of the *C. concisus* isolates. However, *C. concisus* isolated from healthy individuals were found to induce epithelial DNA fragmentation (9 out of 14) and this was correlated with an increase in host cell metabolic activity.

1.4.2.2 Hemolysin and Phospholipases

In Australia, 19 *C. concisus* strains were isolated from patients with diarrhoea in the Royal Children’s Hospital, Melbourne (Russell, 1995), and they were found to have hemolytic activity associated with lysis of mammalian erythrocytes (Istivan et al., 1998). Both secreted and cell-bound haemolytic activities were found in cellular extractions from these clinical strains, indicating the presence of an important virulence factor of *C. concisus*. The cell bound hemolytic activity was related to PLA₂, an outer membrane
phospholipase A (OMPLA) encoded by the pldA gene (Istivan et al., 2004). It was suggested that OMPLA in *C. concisus* is a potential virulence factor associated with tissue destruction during infection. Calcium ion (Ca$^{2+}$) was also shown to impact the stability of the PLA$_2$, both ferric and ferrous ions (Fe$^{3+}$, Fe$^{2+}$) were shown to decreased hemolytic activity (Istivan et al., 2004). Later, Kaakoush et al. (2010) identified a potential hemolysin, TlyA within the genome of *C. concisus* 13826, isolated from a patient with acute gastroenteritis. Kalischuk and Inglis (2011) also showed that *C. concisus* isolates had haemolytic activity against sheep red blood cells.

**1.4.2.3 Exotoxin 9**

Exotoxin 9 was first identified in a plasmid in *C. concisus* UNSWCD (Kaakoush et al., 2011b) and investigated thoroughly. It is a putative virulent determinant in *C. concisus* with high homology with exotoxins found in Gram positive bacteria and may give the ability to survive within host cells. It was detected in highly invasive *C. concisus* strains but not in non-invasive strains by PCR (Kaakoush et al., 2011b). It was suggested the conserved synteny of exotoxin 9 with three other proteins which are a restriction endonuclease, recombinase and DNA methyltransferase has a combined function within the highly invasive *C. concisus* (Deshpande et al., 2013). *C. concisus* and exotoxin 9 levels have been detected by real-time PCR in faecal samples of CD patients and healthy controls (Kaakoush et al., 2014). Both *C. concisus* and exotoxin 9 levels were significantly higher in CD patients than in healthy controls. The authors also reported the correlations of *C. concisus* levels to the intestinal microbiota *Eubacterium, Subdoligranulum* and *Blautia*, while exotoxin 9 levels correlated with *Dialister, Oscillospira, Lachnospira* and *Prevotella*. The correlation of *C. concisus* levels and exotoxin 9 in CD patients suggested exotoxin 9 as a virulence factor in *C. concisus*. Recently, exotoxin 9 was also associated
with acute gastroenteritis where a significantly higher level of this toxin was detected along with *C. concisus* DNA in stool samples by quantitative real-time PCR (Underwood *et al.*, 2015).

1.4.2.4 *Zonula Occludens Toxin (Zot)*

The Zot toxin, first discovered in *Vibrio cholerae*, increases intestinal permeability by targeting the tight junctions and associated with mild to moderate diarrhoea (Fasano *et al.*, 1991). It was identified in *C. concisus* 13826 secretome in 2010 (Kaakoush *et al.*, 2010) and detected in only one out of eight *C. concisus* strains by PCR in 2011 (Kaakoush *et al.*, 2011a). However, whole genome comparisons by Deshpande *et al.* (2013) showed the *zot* was present in another strain of these eight sequences, indicating the primers designed for the *zot* and used in the previous study were not compatible for all *C. concisus* strain. *zot* was detected in 42.8% (6 out of 14) of *C. concisus* isolates from stool samples of both healthy and diarrheic patients (Kalischuk and Inglis, 2011). Very recently *zot* was detected in *C. concisus* isolated from stool samples collected from patients with acute gastroenteritis using quantitative real-time PCR (Underwood *et al.*, 2015). Mahendran *et al.* (2013) was the first to detect *zot* from oral *C. concisus* isolates from patients with active IBD and from healthy control with no significant difference but they have found polymorphisms (*zot*\textsuperscript{808T}, *zot*\textsuperscript{350-351AC} and *zot*\textsuperscript{Multiple}) only in *zot* in *C. concisus* from IBD patients, which resulted in substitution of valine to leucine at position 270 and suggested that polymorphism of *zot* may play a role in virulence. Moreover, Zhang *et al.* (2014) suggested that some oral *C. concisus* strains acquired the *zot* from a virus (a filamentous prophage). *C. concisus* Zot shares conserved motifs with both *V. cholerae* Zot receptor binding domain and human zonulin receptor binding domain. Both *V. cholerae* Zot and
human intestinal Zot analogue are known to increase intestinal permeability by affecting the tight junctions which is also a characteristic feature of IBD (Zhang et al., 2014).

1.4.2.5 S-layer RTX

This toxin is a surface-layer protein (S-layer RTX) which belongs to the RTX family (repeats in the structural toxins) synthesized by a diverse group of Gram negative pathogens (Welch, 1991). These pore-forming toxins are also found in Campylobacter spp. and are recognised as an important virulence factor in C. rectus (Braun et al., 1999). Secretome analysis of the C. concisus strain UNSWCD isolated from an intestinal biopsy of a patient with CD identified an S-layer RTX protein (Kaakoush et al., 2010). The S-layer RTX proteins and their genes were detected by PCR in enteric C. concisus strains (Kalischuk and Inglis, 2011). Later, this was also detected in oral C. concisus isolates (Ismail et al., 2012). Recently, the S-layer RTX expression was observed in absence and presence of H2 in an oral C. concisus strain and no significant effect on S-layer RTX expression levels was found (Lee et al., 2014).

1.4.3 Adhesion and invasion

Adhesion and invasion are prerequisites for a pathogen to colonise the host gastrointestinal epithelial cells. C. concisus strains isolated from children with diarrhoea showed 4- to 100-fold higher invasion rate than pathogenic and invasive C. jejuni and C. coli strains on human HEp-2 cell line (Russell and Ward, 1998). However, Musmanno et al. (1998) have tested C. concisus along with C. jejuni ss. Jejuni, C. jejuni ss. Doylei, C. coli and C. upsaliensis isolated from children with enteritis for adhesion and invasion on the intestinal cell line INT 407 and reported none of the strains adhered. In another study, C. concisus faecal isolates from diarrheic and healthy controls demonstrated comparable
epithelial adherence, invasion, and translocation abilities to that of *C. jejuni* 81–176 on T84 human colonic epithelial cells. The study suggested that there was no difference in mean adherence, invasion, or translocation between isolates from diarrheic and healthy humans (Kalischuk and Inglis, 2011). In the same year, both oral and faecal *C. concisus* isolates were found to invade human HT-29/B6 colon cells and impaired epithelial barrier function (Nielsen *et al.*, 2011). The authors suggested that epithelial barrier dysfunction by oral and faecal *C. concisus* strains could mainly be assigned to apoptotic leaks together with moderate changes to tight junctions.

The invasion characteristics of *C. concisus* strains were investigated by the gentamicin protection assay and the *C. concisus* CD isolate ‘UNSWCD’ was found to be significantly more invasive than other *C. concisus* strains isolated from patients with acute enteritis and from a healthy control (Man *et al.*, 2010a). Scanning electron microscopy showed *C. concisus* UNSWCD used the polar flagellum to adhere intimately to the microvilli (Figure 1.5).

It has been demonstrated that *C. concisus* UNSWCD preferentially attached to intercellular junctional spaces *in vitro* and significantly increased ability to invade TNF-α or IFN-γ -treated Caco-2 cells, suggesting that inflammation (mediated by epithelial cells, monocytes, and macrophages in response to the pathogen) increases the invasive ability of *C. concisus* (Man *et al.*, 2010a). A putative invasin A (InvA) and an outer membrane fibronectin binding protein CadF (involved in adhesion to the host cell) were identified analysing the secretome of *C. concisus* (Kaakoush *et al.*, 2010). Analysis of *C. concisus* UNSWCD genome revealed the presence of an O-antigen ligase that may play a role in the aggregation and adherence on to host cells (Kaakoush *et al.*, 2011a).
Figure 1.5: Host attachment and invasion by *Campylobacter concisus* UNSWCD. A and D, In Caco-2 cells, the polar flagellum of *C. concisus* UNSWCD mediated attachment to the microvillus tip (triangles). B, The flagellum appears to fold around the microvillus (triangle). C, *C. concisus* induced a membrane ruffling-like effect (*). D, *C. concisus* is observed half internalized in the host cell, resulting in a surface protrusion on the host cell membrane, and the flagellated half remains externally exposed. A host cell infected with multiple bacteria displays cell membrane irregularities and uneven texture because of bacteria-induced protrusions (arrows), adapted from Man et al. (2010a).
Chapter 1: General Introduction

The presence of NhaC (sodium-hydrogen antiporter) was linked to *C. concisus* strains with higher adherence and flagella movement and may play a role in virulence (Deshpande *et al.*, 2013). The same group also found that the invasion percentage of *C. concisus* strains isolated from chronic intestinal diseases were > 500-fold higher than those isolated from acute intestinal diseases and from healthy subjects (Kaakoush *et al.*, 2011b). A virulence related plasmid was detected in highly invasive *C. concisus* strains isolated from chronic intestinal diseases, indicating that it may play a role in the invasive potential (Kaakoush *et al.*, 2011b). Invasion assay on oral isolates from IBD patients showed that 50% of them were invasive in Caco2 cells but not the controls (Ismail *et al.*, 2012). *C. concisus* was reported to adhere and invade intestinal epithelial cells, and adherence levels significantly decreased with increasing viscosity of the mucus with no significant change in invasion. It was also shown that *C. concisus* UNSWCD adhere to and invade mucus-producing intestinal cells and adhesion and invasion increased upon accumulation of mucus (Lavrencic *et al.*, 2012). *C. concisus* isolated from porcine samples have been shown via PCR/sequencing to encode CadF (intestinal epithelial cell adhesion protein) in 7 out of 9 (83%) isolates (Scanlon *et al.*, 2013). Recently, Kaakoush *et al.* (2015b) suggested that *C. concisus* can be divided on the basis of specific pathogenic properties within different isolates into pathotypes including adherent invasive *C. concisus* (AICC) and adherent and toxigenic *C. concisus* (AToCC). In this study, comprehensive transcriptome and methylation analysis were performed on intestinal epithelial Caco-2 cells and THP-1 derived macrophages infected with AICC and AToCC strains using RNA-seq and methylation array. The authors found significant difference in these two pathotypes and they suggested AToCC could cause acute gastroenteritis while AICC has been associated with colorectal carcinogenesis and chronic infections.
1.5 Antibiotic resistance

There are several reports on susceptibility testing and on resistance patterns of *C. concisus*. Disc diffusion testing is not reliable for this slow growing organism and the dilution antimicrobial test is too cumbersome. Unfortunately, there are no interpretive data and breakpoints for determining resistance and susceptibility of *C. concisus* in the Clinical and Laboratory Standards Institute (CLSI) approved standards.

In 1981, after identifying the new species, Tanner *et al.* (1981) tested resistance of six *C. concisus* isolates against 17 antimicrobials and found them to be resistant to bacitracin, nalidixic acid, rifampin, and vancomycin. However, Johnson *et al.* (1986) found *C. concisus* strains to be resistant to gentamicin only. Later it was reported that 1 µg/ml minocycline was effective against *C. concisus*. However there was an increase in the *C. concisus* MICs after 6 to 7 weeks exposure to sub-lethal concentrations of minocycline (O'Connor *et al.*, 1990). Furthermore, Engberg *et al.* (2005) reported that *C. concisus* isolates (n = 43) were sensitive to erythromycin, tetracycline, nalidixic acid, streptomycin, ciprofloxacin, gentamicin, colistin, chloramphenicol, sulfamethizole, neomycin, and ampicillin. In 2006, a reduced MIC to ciprofloxacin, hence potential treatment failure, was reported (Vandenberg *et al.*, 2006). In another study, *C. concisus* isolates collected between 1998 and 2006 were reported to have increased resistance to nalidixic acid, ciprofloxacin and erythromycin (Moore *et al.*, 2006). In 2013, 100% resistance to erythromycin was reported in *C. concisus* isolated from porcine samples which may affect the drug of choice for treatment of severe gastroenteritic *Campylobacter* infections (Scanlon *et al.*, 2013). Very recently, *C. concisus* along with other *Campylobacter* spp. were tested by the serum resistance assay and found that killing time depends on concentration of the bacteria and serum; *C. concisus* was reduced to less than
50% of inoculum after 30 min while *C. jejuni* was sensitive and *C. fetus* was completely resistant to serum (Kirk *et al*., 2015).

Antibiotics may influence the course of IBD by decreasing concentrations of bacteria in the gut lumen and altering the composition of intestinal microbiota. In clinical trials for the treatment of IBD different antibiotics including ciprofloxacin, metronidazole, the combination of both, rifaximin, and anti-tuberculous regimens have been evaluated. Rifamixin has shown promising results. Treatment of CD with abscesses or fistulas most often involves ciprofloxacin, metronidazole, or combinations of both. Very few data are available for ulcerative colitis, and mostly consists of small trials evaluating ciprofloxacin, metronidazole and rifaximin. Most trials did not show a benefit for the treatment of active ulcerative colitis with antibiotics. The downsides of antibiotic treatment, especially with recurrent or prolonged courses such as used in IBD, are significant side effects that often cause intolerance to treatment, *C. difficile* infection, and increasing antibiotic resistance (Nitzan *et al*., 2016). Probiotics were found to be ineffective in preventing clinical and/or endoscopic recurrence. By contrast, a combination of a probiotic agent (eg, *Saccharomyces boulardii*) with standard pharmacological therapy were found to promote clinical benefit (Guslandi, 2015).

### 1.6 Biofilm formation by *Campylobacter* spp.

Biofilms are sessile microbial communities in which cells are attached to a surface or air–liquid interface, enveloped within an extracellular polymeric matrix (Costerton *et al*., 1999). The ability to form biofilms is a survival mechanism that microorganisms employ to resist physical, environmental and biological stresses, and to persist in a diverse range of ecological niches. There are some reports of *Campylobacter* spp., mainly *C.*
jejuni forming biofilms (Joshua et al., 2006; Reeser et al., 2007; Gunther and Chen, 2009). Joshua et al. (2006) showed that C. jejuni can exist in three forms in biofilm, those are: i) attached to the glass surface, ii) unattached aggregates (flocs) and iii) pellicles at the liquid-gas interface. In the same year, to understand the cellular mechanisms of biofilm and planktonic C. jejuni, Sampathkumar et al. (2006) studied transcriptional and translational expression profiles. This study showed that C. jejuni goes through some regulation process in biofilms and down regulates metabolism, motility and protein synthesis capabilities and emphasis on iron uptake, oxidative stress defence and membrane transport. The proteomic analysis of C. jejuni 11168 biofilm showed higher levels of expression of proteins involved in the motility complex, including the flagellins (FlaA, FlaB), the filament cap (FliD), the basal body (FlgG, FlgG2), and the chemotactic protein (CheA) in biofilms. Different genes involved in motility were also investigated such as flhA, fliA, flaA, flaB, flag, flaC and it was found that inactivation of any of these genes resulted in a delay in biofilm formation in C. jejuni (Kalmokoff et al., 2006). Flagella and quorum-sensing (QS) play important roles in biofilm formation. Inactivation of flaAB (flagella subunits) and luxS (responsible for QS) resulted in a reduced biofilm formation in C. jejuni. Hence it was suggested that both flab and luxS are required for biofilm formation (Reeser et al., 2007). Since then, many researchers have investigated luxS in C. jejuni and demonstrated that this gene is involved in a variety of physiologic pathways, motility, autoagglutination, cytolethal distending toxin (CDT) expression, flagellar expression, oxidative stress, and animal colonization (Jeon et al., 2005; Reeser et al., 2007; Quinones et al., 2009; Plummer et al., 2011). However, the involvement of LuxS in the above mentioned characteristics in C. concisus has not been well studied yet. Hence, the discussion for the role of LuxS on Campylobacter spp. described in the following sections is mainly focused on C. jejuni.
In 2009, fourteen *Campylobacter* species, both microaerophilic and hydrogen-requiring microaerophilic were tested on different surfaces such as glass, stainless steel, and polystyrene plastic for biofilm-forming ability. Of the eight microaerophilic *Campylobacter* species tested, *C. jejuni*, *C. coli*, *C. lari*, *C. upsaliensis*, *C. sputorum*, *C. hyointestinalis* and *C. helveticus C. fetus*, only *C. jejuni* strain 81–176 reliably produced a visible biofilm on multiple surfaces. While, all six strains of the hydrogen-requiring microaerophilic *Campylobacter* species *C. rectus*, *C. showae*, *C. mucosalis*, *C. concisus*, *C. curvus* and *C. gracilis* reliably produced visible biofilms on multiple surfaces. The biofilms were also visualised by electron microscopy which found relatively homogeneous biofilms produced by the hydrogen-requiring microaerophilic strains compared to the microaerophilics (Gunther and Chen, 2009). Micro-colonies were detected in the form of mucosal biofilm from biopsy samples from patients with Barrett’s Esophagus (BE). *C. concisus* and *C. rectus* were detected in 4 (57%) of 7 patients with BE but not in the control subjects (Macfarlane et al., 2007). This study has been supported recently by Blackett et al. (2013), where they found an oesophageal biofilm in patients with Barrett's oesophagus (BO) and gastro-oesophageal reflux disease (GERD). Interestingly, *C. concisus* was the dominant species and the authors suggested emergence of *C. concisus* as the dominant new colonist in the refluxed oesophagus. Moreover, *C. concisus* isolated from IBD patients, gastroenteritis and healthy individuals were reported as equally capable of forming biofilm on glass cover slip and not related to motility suggesting that biofilm formation is a general strategy used by *C. concisus* to colonise the natural environment (Lavrencic et al., 2012). *C. concisus* has also been isolated from subgingival microbiota of individuals with HIV (Pereira et al., 2014) and human dental plaque biofilm (Marchesan et al., 2015). The ability hydrogen-requiring microaerophilic of *Campylobacter* spp. forming more biofilm than the microaerophilics (Gunther and Chen, 2009) may explain the
survival of *C. concisus* in the oral cavity. However to date, there has been no thorough investigation of *C. concisus* biofilm formation.

### 1.6.1 Mechanism of biofilm formation

Biofilm formation is commonly considered to occur in three main stages (Joo and Otto, 2012). Recent research has provided molecular insight into all three stages shown in Figure 1.6 (Adler *et al.*, 2014):

1. **Attachment:** the individual adherent cells that initiate biofilm formation on a surface are surrounded by only small amounts of exopolymeric materials and many are capable of independent movement by means of a pilus-mediated twitching or gliding. In this stage cells may leave the surface to resume the planktonic form.

2. **Proliferation and maturation:** the cells attach irreversibly, a step mediated mainly by exopolymeric substances or extracellular matrix. The cells lose their flagella-driven motility; leading to formation of the characteristic, mature biofilm structure. The function of the matrix is to provide adhesion between bacterial cells, hence to form a multilayered biofilm. It contains exopolysaccharides, proteins, extracellular DNA (eDNA) and other polymers. These components not only help to develop the fully mature biofilms, but also provide protection from antibiotics and host defences. As biofilms mature they develop the basic micro colonies or water channel architecture.

3. **Detachment:** which is also often called dispersal. Individual microcolonies may detach from the surface or may give rise to planktonic revertants that swim or float away from the matrix-enclosed structures, leaving hollow remnants of microcolonies or empty spaces that become parts of the water channels.
Figure 1.6: Phases of *in-vivo* biofilm development. Biofilms develop via initial attachment, which depends on transport of the bacteria to a surface, which is passive in the case of non-motile bacteria such as staphylococci (yellow), and active in the case of motile bacteria such as *P. aeruginosa* (red). Attachment itself is governed by specific protein-protein interactions of bacterial surface with human matrix proteins. Attachment to an abiotic surface such as a catheter depends on bacterial surface hydrophobicity, but this mechanism is believed to have minor importance in vivo. Subsequent steps do not differ in principle between motile and non-motile bacteria. They involve proliferation, embedding in an extracellular matrix, and maturation. The latter depends on cell-cell disruptive factors, recently identified to be primarily surfactants. Strong production of surfactants, which are controlled by QS, leads to biofilm detachment (dispersal). In the case of motile bacteria, up-regulation of motility, starting in the center of biofilm “mushroom caps” assists dispersal, adapted from Joo and Otto (2012).
The proteome of biofilm changes continuously during development. Radical changes in the protein profile of *P. aeruginosa* have been demonstrated from planktonic bacterial growth; over 800 proteins were shown to have six-fold or greater difference in expression (Sauer *et al.*, 2002). The genetic component important in biofilm formation is explained in Figure 1.7.

### 1.6.2 Cell to cell signalling (Quorum-sensing)

In 1983, the first evidence of bacterial communication came from a study of density-dependent bioluminescence of marine symbiotic *Vibrio fischeri* (Engebrecht *et al.*, 1983). Small diffusible signalling compounds produced by the bacteria while they are growing named autoinducers (AIs) were found to control the induction of bioluminescence.

The study conducted by Engebrecht *et al.* (1983) explains the quorum-sensing model in Gram negative bacteria where acyl homoserine lactone (HSL) or autoinducer (AI-1) produced by the bacteria accumulates in the extracellular environment and freely diffuses into the bacterial cytoplasm. When intracellular AI levels reach a critical threshold concentration, a signal transduction cascade is initiated and it binds to a cellular transcriptional activator (*luxR* homolog) and induces the luciferase gene to illuminate (Bassler, 1999). The author also demonstrated a similar system in over 30 species of Gram negative bacteria where it controls over a variety of cell density-dependent processes, while Gram positive species have been demonstrated to produce small peptide signalling molecules that are secreted through the action of an ABC (ATP-binding cassette) transporter protein system (Bassler, 1999).
Figure 1.7: Summary of genetic components important in biofilm formation. Not all bacteria require all of these factors, and distinct genetic pathways can be utilized by a single species to form a biofilm, adapted from Pratt and Kolter (1999).
The extracellular AIs are then recognized by a cognate two-component sensor kinase that initiates a phosphorelay signal transduction cascade.

Interestingly, there is a third quorum-sensing mechanism that appears to be highly conserved over a variety of Gram negative and Gram positive organisms (Federle and Bassler, 2003). In these organisms the luxS gene is responsible for the synthesis of a novel AI termed autoinducer-2 (AI-2), which has mechanisms reminiscent of both the Gram positive and Gram negative systems (Figure 1.8). The model system, Vibrio harveyi, uses a well described two component sensor kinase system (LuxP/Q) to initiate a phosphorelay signal transduction event that ultimately leads to dephosphorylation of the response regulator (LuxO) and changes in gene expression of the luciferase structural operon luxCDABE (Bassler et al., 1993). However, some Gram negative bacteria like E. coli have demonstrated dependency on a periplasmic binding protein (LsrB) that recognizes the AI-2 and enter via an ABC transporter (encoded by the lsrACDBFGE operon) where the AI-2 is phosphorylated and regulates gene expression via LsrR (Xavier and Bassler, 2005).

Furthermore from the C. concisus 13826 genome published in NCBI (https://www.ncbi.nlm.nih.gov/nuccore/CP000792), it has been found that C. concisus 13826 (Accession no CP000792) does not possess acyl-HSL synthetases and the AI-1 sensing system, but does possesses a luxS homolog of with 72% identity to gene sequence of V. fischeri.
Figure 1.8: AI-2 signalling molecules and receptors. (a) The LuxS product DPD spontaneously undergoes cyclization and hydration reactions in solution to form R-THMF (detected by Salmonella Typhimurium) and S-THMF. In the presence of boron, the latter molecule further reacts to form S-THMF-borate (detected by V. harveyi). (b) These AI-2 ligands are coordinated by different interactions (hydrogen bonds or salt bridges, dashed yellow lines) with different amino acids residues in the binding sites of their respective receptors: LuxP containing S-THMF-borate from V. harveyi, and LsrB containing R-THMF from S. Typhimurium, although the structure of the receptors (c), shown by ribbon diagrams coloured in rainbow order from the N- to the C-terminus, reveals much similarity between these two proteins, adapted from Pereira et al. (2013).
1.7 Role of luxS in QS

The production of AI-2 by S-ribosyl homocysteinase or LuxS was first described in *V. harveyi* (Bassler, 1999). LuxS cleaves the thio-ether bond in S-ribosylhomocysteine (SRH) to produce homocysteine and 4,5-dihydroxy-2,3-pentanedione (DPD) which is then spontaneously cyclized to form AI-2 (De Keersmaecker et al., 2005). While the bacterial population is growing, it produces soluble AI-2 and when AI-2 environmental levels reach a critical concentration it gives a signal for bioluminescence of the organism (Pereira et al., 2013). As this system has been detected in a number of species it has become clear that the AI-2 molecules in different species are not necessarily identical, but are most likely derived from the same precursor. So, the name AI-2 is used collectively to represent a variety of molecular variants that result from spontaneous changes to DPD and act as ligands of AI-2 receptors (Miller et al., 2004). It was suggested that these molecules are able to interconvert allowing for the inclusion of mixed bacterial populations in the quorum recognized by a single species (Xavier and Bassler, 2005). The molecular form of AI-2 produced by *C. jejuni* is still unknown. However, there are few reports about its activity as a luxP ligand during in vitro bioluminescence assays with a *V. harveyi* reporter strain (Elvers and Park, 2002; Plummer et al., 2011). To our knowledge, no research has been done to investigate the influence of luxS in *C. concisus*.

1.7.1 The structure and activity of LuxS

In *Helicobacter pylori*, *Deinococcus radiodurans*, and *Haemophilus influenzae*, structural studies of LuxS have shown that it forms a secondary structure containing a four-stranded antiparallel β sheet interacting with four α helices (Lewis et al., 2001). This structural fold observed in all of these crystals is highly conserved and appears to be a novel fold in the alpha-beta family. At the interface of the homodimer two identical active
sites are present with a tetrahedrally coordinated Fe$^{2+}$ (Hilgers and Ludwig, 2001). There is also evidence of two deep pockets adjacent to the active sites. The function of the pockets is still not clear, but they hold a covalently bound homocysteine molecule (Ruzheinikov et al., 2001). The proposed enzymatic activity involves a series of proton transfer reactions, which are catalysed by a Fe$^{2+}$ ion and two general acids/bases in the LuxS active site, resulting in the migration of the ribose carbonyl group from its C1 to C3 position (Zhu et al., 2004). Two universally conserved residues, Cys-84 and Glu-57 (aa numbering in E. coli LuxS), have been shown to act as critical general acids/bases during catalysis (Rajan et al., 2005; Zhu et al., 2004). In C. jejuni a naturally occurring G92D mutation was identified in some clones of C. jejuni 81116 (Plummer et al., 2011). The glycine at the 92 position is invariant in active enzymes described to date. Strains with this mutation fail to produce significant levels of AI-2 in the biological systems utilized. This finding underlines the significance of fully characterizing isolates used for quorum-sensing research even if they are strains previously reported in the literature to have functional LuxS systems.

LuxS of C. concisus has not been crystallized; however, the gene sequence shows a high degree of homology to other sequenced strains (70% to V. harveyi LuxS, 71% to E. coli LuxS, and 70% to H. influenzae LuxS). The crystal structure of H. influenzae was determined by Lewis et al. (2001) (Figure 1.9), retrieved from http://www.rcsb.org/pdb/home/home.do.
Figure 1.9: Crystal structure of *Haemophilus influenzae* LuxS, (70% aa identity to LuxS from *C. concisus*) with two ligand methionine and zinc ion retrieved from Lewis *et al.* (2001), (http://www.rcsb.org/pdb/explore/explore.do?structureId=1J6W).
1.7.2 Metabolic role of LuxS

The primary function of LuxS in bacterial cells is within the s-adenosyl-homocysteine recycling pathway (Figure 1.10). Functionally, LuxS is the enzyme called s-adenosyl-homocysteinase responsible for hydrolysis of s-adenosylhomocysteine to homocysteine which is then further metabolized to reclaim s-adenosylmethionine (SAM) (Winzer et al., 2002). As a consequence, mutagenesis of LuxS results in a loss of normal SAM recycling and therefore loss of AI-2 synthesis. This fact has greatly hindered the ability to specifically attribute any phenotypic differences associated with mutagenesis of LuxS with a specific loss of quorum-sensing ability. Mutagenesis of LuxS results in alteration of the extracellular concentration of other SAM pathway metabolites. Plummer et al. (2011) showed that the extracellular concentration of s-ribosylhomocysteine increases significantly in the culture supernatant of strains deficient in normal LuxS activity by using a recombinant LuxS protein bioassay. Based on these findings, it is likely that many of the SAM pathway intermediates are present in altered concentrations both intracellularly and extracellularly in ΔluxS mutant strains. The SAM pathway is the primary means of methyl recycling in bacteria (Parveen and Cornell, 2011). Methyl groups donated by SAM are critical for the activity of methyl transferases that are intimately associated with bacterial DNA methylation, chemotaxis, motility, and a variety of other metabolic and biosynthetic reactions. SAM is also critical in bacterial polyamine formation and vitamin synthesis (Parveen and Cornell, 2011). As a consequence, alteration of SAM recycling associated with luxS mutagenesis can have significant impacts on bacterial metabolism.
Figure 1.10: The AMC (activated methyl cycle) and synthesis of autoinducers. (a) Transfer of a methyl group from SAM to an acceptor molecule generates the toxic intermediate, SAH. (b) Organisms that contain the SahH enzyme convert SAH directly to homocysteine. (c) Those with LuxS catalyse this reaction in two steps: Pfs first generates SRH and adenine from SAH, then LuxS converts SRH into homocysteine and DPD, the linear form of AI-2. (d) Homocysteine is regenerated into SAM via a series of reactions involving methionine. THF-CH3 and THPG stand for N-methyltetrahydrofolate and 5-methyltetrahydropteroyl glutamate, respectively. (e) LuxI-type proteins also use SAM and acyl carrier proteins (Acyl-ACP) as substrates for the production of AHL autoinducers. (f) *Vibrio cholerae* autoinducer CAI-1 is also synthesized from SAM and decanoyl Coenzyme A via a multistep reaction involving CqsA, a pyridoxal phosphatase (PLP)-aminotransferase like enzyme, and VC1059. *V. harveyi* has a homologue to CqsA that uses octanoyl CoA yielding Z-3-aminoundec-2-en-4-one, adapted from Pereira et al. (2013).
1.7.3 LuxS and growth

The effect of LuxS on the growth rate of enterohemorrhagic *E. coli* (EHEC) O157:H7 was studied by hybridization of a gene array with cDNA synthesized from RNA that was extracted from EHEC strain 86-24 and its isogenic Δ*luxS* mutant (Sperandio *et al.*, 2001). It was observed that 235 genes were up-regulated and 169 were down-regulated in the wild type strain including several involved in cell division, as well as ribosomal and tRNA genes compared to in the Δ*luxS* mutant. As a result, the Δ*luxS* mutant grew faster than the wild type strain (generation times of 37.5 and 60 min, respectively). At the same time a *Streptococcus pyogenes* Δ*luxS* mutant strain demonstrated a 30% reduction in its growth rate in Todd–Hewitt yeast extract media (Lyon *et al.*, 2001). However in *E. coli* K-12 strain W3110, the growth rates of the Δ*luxS* mutant and its parent strain were indistinguishable in LB or in LB and glucose at 30°C (Wang *et al.*, 2005). The growth rate was compared with *C. jejuni* NCTC 11168 strain and Δ*luxS* mutants and no significant reduction was found in the mutant strain (Elvers and Park, 2002). Other studies have shown similar results for equal growth rates between the wild type and Δ*luxS* mutant strains of *C. jejuni* 81116 (Jeon *et al.*, 2003), M129 (Reeser *et al.*, 2007), and NCTC 11168 (Holmes *et al.*, 2009). However, *C. jejuni* 81–176 Δ*luxS* mutant showed significantly longer doubling time compared to the wild type strain when grown at 37°C, but not at 42°C (He *et al.*, 2008). Later it was demonstrated that the Δ*luxS* mutant of the 81–176 strain had a decreased growth rate during the exponential phase but reached stationary phase at the same time as its wild type counterpart when grown at 42°C (Quinones *et al.*, 2009). Based on these collective results it appears that the growth rate reduction of *C. jejuni* depends on the strain type and complementation of *luxS* is a necessary step to confirm the effect in growth reduction. There is no report on the effect of LuxS on the growth rate of *C. concisus*. 
1.7.4 LuxS and biofilm formation

Quorum-sensing via AI-2 plays a significant role in oral biofilm formation. The role of AI-2 was investigated on *Streptococcus mutans* biofilm formation by constructing a *luxS*-null mutant (Yoshida *et al.*, 2005). Biofilm formation by the Δ*luxS* mutant was found to be significantly reduced compared to the wild type, however, biofilms of the Δ*luxS* mutant were shown to form larger clumps compared to the parental strain. Interestingly, the biofilm defect of the Δ*luxS* mutant of *S. mutans* was complemented by other oral bacteria such as *S. gordonii*, *S. sobrinus*, and *S. anginosus* (Yoshida *et al.*, 2005). Disruption of *luxS* in *S. anginosus* resulted in deficient biofilm formation, but no effect on planktonic growth rate of *S. anginosus* (Petersen *et al.*, 2006). In the following year, Reeser *et al.* (2007) demonstrated that *luxS* and flagellar structure may be important for *C. jejuni* biofilm growth since isolates that were mutagenized in either the flagellar proteins or *luxS* had significantly decreased biofilm formation using a simple static biofilm assay viz. the crystal violet assay. Using *C. jejuni* strain M129, a human clinical isolate, it was demonstrated that the Δ*luxS* mutant had a statistically significant reduction in crystal violet staining which indicated a reduction in biofilm formation following a 48h and 72h incubation at 37°C in a 10% CO₂ environment. Similar reductions in biofilm formation were observed in a Δ*flaAB* mutant, but it was unclear if the changes in biofilm formation were a direct response to the loss of a functional LuxS enzyme or secondary to altered transcription of *flaA* (Jeon *et al.*, 2003). However, the overall significance of biofilms in the physiology and survival of *Campylobacter* spp. is still not well understood.

1.7.5 LuxS and chemotaxis

EHEC O157:H7 strain 86-24 and its isogenic Δ*luxS* mutant were investigated by a gene array and the expression of numerous genes (*cheW, cheA, tar, tapE, cheR, cheY* and
cheZ was found to be involved in chemotaxis and motility, which were both up-regulated in the wild type strain compared to in the ΔluxS mutant (Sperandio et al., 2001). Unlike E. coli, Campylobacter and Helicobacter species regulate their motility by chemotactic signalling systems (Lertsethtakarn et al., 2011). For chemotactic signal transduction, four different groups of proteins are necessary: chemoreceptors, core signal proteins, accessory proteins and flagellar switch proteins (Lertsethtakarn et al., 2011). However, the expression of neither cheA, cheW (core signal proteins) nor cheB, cheR and cheV (accessory proteins) were differentially regulated in the C. jejuni 81-176 luxS mutant grown at 42°C (He et al., 2008). Later, a down-regulation of cheA in the ΔluxS mutant of C. jejuni NCTC 11168 was described (Holmes et al., 2009). Quinones et al. (2009) used the hard-agar plug method to test the chemotactic response of C. jejuni 81–176 to both amino acids and organic acids and found an enhanced chemo-attraction towards amino acids (aspartate, asparagine, glutamate, and glutamine) and reduced chemo-attraction towards organic acids for the C. jejuni 81-176 luxS mutant as compared to the wild type strain. The authors suggested that these changes might be a response to altered metabolic activity of the ΔluxS mutant that resulted in disrupted amino acid metabolism and carbon compound catabolism and may contribute to decrease the chicken colonization which was fully covered by complementation. On the other hand, a ΔluxS mutant of H. pylori strain G27 showed reduced stop frequency (cessation of movement in any direction for three or more frames by phase-contrast light microscope) in liquid media which was restored by the addition of AI-2 or DPD (Rader et al., 2011).

1.7.6 LuxS and motility

The effects of AI-2 were studied on various phenotypes associated with EHEC infections and it was found that EHEC is attracted by AI-2 in agarose plug chemotaxis
assays, resulting in an increased swimming motility (Bansal et al., 2008). A *H. pylori* Δ*luxS* mutant demonstrated a decreased motility on soft agar plates when compared to the wild type, which could be complemented by a wild type copy of the *luxS* gene and by the addition of cell-free supernatant containing AI-2 (Rader et al., 2007). Later, the same group reported that deletion of the *luxS* resulted in swimming behaviour with a reduced frequency of stops as compared to the wild type strain. Stopping frequency was restored to wild type levels by genetic complementation or by chemical complementation by addition of synthetic DPD. It was suggested that *H. pylori* recognizes the LuxS-produced AI-2 as a chemorepellent via the chemoreceptor TlpB (Rader et al., 2011).

The role of LuxS in the motility of *C. jejuni* was first demonstrated by using a semisolid medium where the Δ*luxS* mutant formed a statistically smaller halo compared to the wild type strain (Elvers and Park, 2002). Subsequently, many studies were in agreement that motility on semisolid media is decreased in Δ*luxS* mutant strains (Jeon et al., 2003; Holmes et al., 2009; Quinones et al., 2009). The motility was partially restored by genetic complementation of the *luxS* gene (Quinones et al., 2009; Plummer et al., 2011). However, Jeon et al. (2003) attempted to explain the decreased motility of Δ*luxS* mutants by phase contrast microscopy on the wild type and Δ*luxS* mutant strains but failed to observe any differences in motility speed or motility pattern. Then to investigate the molecular basis for the motility phenotype they evaluated expression of the two major flagellin genes (*flaA* and *flaB*) and observed approximately 43% reduction in the transcription of *flaA* in the Δ*luxS* mutant background compared to wild type by the primer extension assay. However, no difference was observed in FlaA expression levels using western blotting nor flagellar structure based on transmission electron microscopy. Later, a
down regulation of *flaA* (4.2-fold down-regulated in Δ*luxS* mutant compared to wild type) was confirmed using cDNA microarray (Holmes *et al.*, 2009).

### 1.7.7 LuxS and cellular adherence and invasion

EHEC attachment to HeLa cells was enhanced by AI-2 exposure. Further investigation at the transcriptome level suggested that exposure to AI-2 altered the expression of 23 LEE (locus of enterocyte effacement, which is a chromosomal pathogenicity island of size 35 kb) genes directly involved in the production of virulence determinants, as well as other genes associated with virulence in a temporally defined manner (Bansal *et al.*, 2008). However, *S. enterica* serovar Typhimurium was investigated to determine whether the *luxS* gene had any effect to the ability to invade epithelial cells. No differences were found between the wild type strain and its isogenic Δ*luxS* mutant with their ability to invade epithelial cells (Perrett *et al.*, 2009).

In *C. jejuni*, the impact of *luxS* mutagenesis is determined by both the strain being tested and the eukaryotic cell type used for the assay. Elvers and Park (2002) failed to detect a difference in adherence and invasion of Caco-2 cells between the wild type and mutant strains. In their study on *C. jejuni* NCTC11168 the authors determined an internalization rate of 0.08% for the wild type and 0.05% for the Δ*luxS* mutant, which was statistically insignificant. Later, the adherence of *C. jejuni* to LMH cells was evaluated and demonstrated a significant reduction (*p* < 0.01) in adherence of the Δ*luxS* mutant of 81–176 when compared to the wild type strain, which was complemented.
1.7.8 LuxS and virulence

Expression of the Locus of Enterocyte Effacement (LEE) operons encoding the type III secretion system, translocated intimin receptor, and intimin is regulated by quorum-sensing in both EHEC and enteropathogenic E. coli (EPEC) (Sperandio et al., 1999). These results suggested that intestinal colonization by E. coli O157:H7, which has an unusually low infectious dose, could be induced by quorum-sensing of signals produced by non-pathogenic E. coli of the normal intestinal microbiota. When the virulence of the oral pathogen S. mutans was investigated it was reported that the high cell density triggered high level mutacin I gene transcription and this response was abolished upon the deletion of luxS. Furthermore, the reduction of both the mutacin I structural gene mutA and the mutacin I transcriptional activator mutR was confirmed in the ΔluxS mutant transcription by real-time reverse transcription polymerase chain reaction (RT-PCR) (Merritt et al., 2005). The disruption of luxS in S. intermedius affected hyaluronidase and intermedilysin gene expression indicating that AI-2 communication is involved in intermedilysin expression (Pecharki et al., 2008). Evidence has shown that in S. enterica serovar Typhimurium, the luxS gene is necessary for Salmonella virulence phenotypes. Transcription assays showed that the cell-density-dependent induction of the invF gene was abolished in a Salmonella strain with the luxS gene deleted (Choi et al., 2007).

There are limited data available in the role of LuxS in the virulence of C. jejuni. Recently, the role of LuxS in C. jejuni was evaluated as a cause of abortion in guinea pigs. Orally inoculated C. jejuni IA3902 strain (highly virulent sheep abortion strain) was demonstrated to cause abortion in pregnant guinea pig while its isogenic ΔluxS mutant completely lost of ability to colonize the intestinal tract also failed to induce abortion (Plummer et al., 2012). The abortion phenotype was completely restored by genetic
complementation. However, when the organism was inoculated into guinea pigs by the intraperitoneal route, no difference in virulence (abortion induction) was observed between the $\Delta luxS$ mutant and the wild type strain, which suggested LuxS plays an important role in the translocation of *C. jejuni* out of the intestine and into the systemic vasculature where it can move to the placenta and induce abortions.

### 1.7.9 LuxS and haemolysis

In *S. pyogenes*, the inactivation of luxS resulted in mutants having aberrant expression of several virulence properties that are regulated in response to growth phase, including enhanced haemolytic activity. Enhanced haemolytic activity of the $\Delta luxS$ mutant was also shown to be linked with an increased expression of the haemolysin S-associated gene *sagA* (Lyon *et al.*, 2001). However, Pecharki *et al.* (2008) has found that disruption of luxS in *S. intermedius* affected hyaluronidase and intermedilysin gene expressions and resulted in five-fold decrease in haemolytic activity of the $\Delta luxS$ mutant. They also showed that the AI-2 precursor DPD complemented the lack of AI-2 production by the mutant and restored haemolytic activity. Later, it was shown in *S. suis* that the luxS gene deletion caused a significant decrease of haemolytic activity, while the haemolytic activity of the mutant can be restored by genetic complementation (Wang *et al.*, 2011). So far there is no report on the disruption of luxS and its effect on haemolysis of *C. jejuni*.

### 1.8 Molecular typing of *C. concisus*

#### 1.8.1 Genetic diversity of *C. concisus*

The first study to explore genetic diversity in *C. concisus* was undertaken by Vandamme *et al.* (1989) that used DNA: DNA hybridization to investigate diarrhoeal isolates and raised the possibility that *C. concisus* was genetically diverse. In this study
phenotypically confirmed *C. concisus* strains shown 42–50% DNA: DNA hybridization values with the type and reference strains of oral origin. In 1996, further evidence for genetic diversity within *C. concisus* isolates of faecal origin was provided where RAPD (Randomly amplified polymorphic DNA) analysis was used to compare *C. concisus* isolated from faecal samples of 37 children with diarrhoea attending the same day care centre. *C. concisus* strains isolated from 35 of the 37 children (94.6%) produced distinct RAPD profiles (Van Etterijck *et al.*, 1996). Based on these results it was concluded that *C. concisus* was genetically diverse and that the species was likely to be “a taxonomic continuum comprised of several genomospecies”. Another study used pulsed field gel electrophoresis (PFGE) with restriction enzyme *NotI* to determine the genetic diversity of 53 *C. concisus* strains isolated from faecal samples of children with diarrhoea and found 51 distinct patterns of macrorestriction fragments out of 53 samples, while 2 strains were resistant to *NotI* digestion (Matsheka *et al.*, 2002). Later, RAPD analysis was performed using (GTG)$_5$ oligonucleotide as primer to differentiate 100 isolates of *C. concisus* strains and 86% of those strains were found to be genetically distinct and rest were classified into five profile groups on the basis of DNA fingerprinting and suggested that this primer could be useful to distinguish *C. concisus* from a large pool of unrelated strains (Matsheka *et al.*, 2006). In 2004, the PCR method for detection of *C. concisus* used by Bastyns *et al.* (1995) was modified and used to group 19 clinical isolates from children with diarrhoea into two genomospecies using primers MUC1/CON1 (genomospecies A) and MUC1/CON2 (genomospecies B) designed from the 23S rDNA region (Istivan *et al.*, 2004).

In 2005, the genotypic characteristics of 39 isolates of *C. concisus* isolates were investigated from patients with diarrhoea, three from healthy individuals and the type
strain CCUG 13144 (Engberg et al., 2005). In this study, C. concisus were grouped according to their cytolethal distending toxin (CDT)-like effect on Vero cells, SDS-PAGE protein profiles and PCR amplification of 23S rDNA, automated ribotyping (RiboPrinting) and RAPD analysis. Two discordance groups were found by protein profiling and PCR amplification of the 23S rRNA gene, automated ribotyping showed 34 of the 43 isolates to have distinct patterns and RAPD showed 37 unique profiles and the authors concluded that C. concisus consists of at least two genomospecies with extensive genetic diversity. In the same year, Aabenhus et al. (2005a) examined 62 C. concisus clinical isolates (56 diarrhoeal strains, four oral strains, and the oral CCUG 13114 and intestinal CCUG 19995 type strains) by amplified fragment length polymorphism (AFLP) profile and all of them gave unique profiles. However, they were able to divide all strains into four distinct clusters. AFLP cluster 1 contained the type strain of oral origin (CCUG13144) as well as 22 other clinical isolates. Cluster 2 contained the reference strain CCUG19995 originally isolated from a human diarrhoea sample, as well as 32 other clinical isolates. AFLP cluster 3 consisted of a single diarrhoeal isolate from an immuno-compromised patient, whereas cluster 4 contained five strains all of which were isolated from severely immune deficient patients, a finding that led the authors to suggest that the strains in AFLP cluster 4 may be less invasive. As result of their study, the authors concluded that C. concisus contains at least four distinct genomospecies that may vary in their pathogenic potential.

Kalischuk and Inglis (2011) compared the genotypes of C. concisus faecal isolates from gastrointestinal patients and asymptomatic healthy individuals by AFLP analysis and genomospecies-specific 23S rRNA gene PCR. Of the 22 isolates examined, six were assigned to genomospecies A and 12 to genomospecies B, while 3 isolates generated PCR products with both the genomospecies A and B primer sets, and were designated
genospecies A/B, while one isolate failed to amplify with either primer set. Based on AFLP analysis the *C. concisus* isolates were shown to cluster into two phylotypes. AFLP cluster1 contained all genospecies A including the type strain LMG7788, four isolates from healthy controls and one isolate from a patient with diarrhoea. AFLP cluster2 had 17 *C. concisus* isolates, 12 belonged to genospecies B, 3 to genospecies A/B, 1 to genospecies A, and the final untypable isolate.

### 1.8.2 Typing of *C. concisus* on the basis of phenotypic characteristics

Based on current taxonomic guidelines, On and Harrington (2000) suggested that diarrhoeal isolates and oral strains should actually be distinct genomic species but phenotypically indistinguishable and *C. concisus* should be regarded as a ‘complex species comprising at least two genotypes’. Later, genetic diversity between oral and faecal isolates of *C. concisus* was confirmed by examining the protein profiles using SDS-PAGE of *C. concisus* strains isolated from patients with a range of symptoms including diarrhoea and dyspepsia. This showed 85% of *C. concisus* faecal strains differed from the oral reference type strain *C. concisus* ATCC 33237^T^ (Aabenhus et al., 2002b). In this study, *C. concisus* isolates were assigned into two broad groups, Group1 which included *C. concisus* strains resembling the reference type strain, and Group2 which comprised *C. concisus* strains whose protein profiles differed from that of the reference type strain. Interestingly, Group2 (83/98, 85%) predominantly comprised *C. concisus* strains isolated from patients with diarrhoea, immunocompetent patients and children.

A lectin typing system from plant source, which used a panel of four lectins, was employed to type 44 *C. concisus* clinical isolates collected from patients with malignancies, HIV, IBD, and other conditions, who had suffered diarrhoea and upper
gastrointestinal dyspepsia, in addition to the type strain ATCC 33237\textsuperscript{T} of oral origin (Aabenhus et al., 2002a). This typing system grouped the 45 isolates into 13 lectin reaction patterns; however, the authors could not correlate the reaction patterns with the clinical category of the patients. Istivan et al. (2004) detected five different protein profiles using SDS-PAGE analysis of OMP extracts for 19 \textit{C. concisus} strains isolated from children with gastroenteritis. Later, these researchers again typed 109 strains of \textit{C. concisus} on the basis of protein profiles, conventional biochemical testing and antibiotic susceptibility and found two distinct groups based on protein profiles (Aabenhus et al., 2005b).

1.9 Genome analysis of \textit{Campylobacter} spp.

Among the genus \textit{Campylobacter}, full genomes of \textit{C. jejuni}, \textit{C. fetus}, \textit{C. concisus}, \textit{C. curvus}, \textit{C. lari}, and \textit{C. hominis} have been sequenced. \textit{C. jejuni} NCTC 11168 was the first genome sequence of \textit{Campylobacter} species published in 2000 and re-annotated in 2007 (Parkhill et al., 2000; Gundogdu et al., 2007). Currently 912 genome assemblies of \textit{C. jejuni} strains are available in the NCBI database (http://www.ncbi.nlm.nih.gov) (inspected on 13.12.2016) and in the Genomes Online Database (GOLD) (https://gold.jgi.doe.gov/), with a total of 1,617 \textit{C. jejuni} genome sequence data entries available (inspected on 13.12.2016). The whole genome assembly of \textit{C. coli} RM4661 is available in the database (Tatusova et al., 2014) along with other 796 \textit{C. coli} sequences in the NCBI database (inspected on 13.12.2016). In addition a total of 1,096 \textit{C. coli} genome sequence data entries are available on the Genomes Online Database (GOLD) (inspected on 13.12.2016). \textit{C. concisus} 13826 and ATCC 33237\textsuperscript{T} are the reference genome sequences available with other seven \textit{C. concisus} species in the database. \textit{Campylobacter} spp. have relatively smaller sized genomes compared to other bacteria. The genome sizes of \textit{C.}
jejuni strains range from 1.63–1.85 Mb, C. coli strains have 1.6–2.1 Mb genomes and the genome size for C. concisus strains is 1.8–2.11 Mb. Based on the full genome sequence data Campylobacter spp. have low %GC contents ranging from of 29.6–39.2. Proteins encoded by Campylobacter spp. range from 1,580 to 2,092. Plasmids were identified in some of the full genome-sequenced Campylobacter spp. strains.

1.9.1 Whole genome sequence of C. concisus

C. concisus 13826 or ATCC BAA-1457 was the first sequenced strain within this species. This reference strain was isolated in Denmark from the faeces of a patient with bloody diarrhoea. The complete genome sequence of this isolate was submitted in NCBI Genome database in September, 2007 (http://www.ncbi.nlm.nih.gov/genome/1439).

The first whole genome shotgun sequencing of C. concisus was performed in 2011 (Deshpande et al., 2011). The whole genome (1.8 Mb) of C. concisus strain UNSWCD, isolated from a patient with CD was sequenced using the ‘Illumina genome analyser’ and it was assembled by the de novo assembly tool ‘Velvet’. When the genome of UNSWCD was compared with the reference genome sequence C. concisus 13826, 1593 genes were found to be conserved across both strains, while 138 genes (7.8%) from UNSWCD and 281(13.98%) from C. concisus 13826 were found to be unique. With a cut off value of 70% identity plus at least 85% gene length coverage, 76% of genes were homologs between C. concisus 13826 and UNSWCD (Kaakoush et al., 2011a). Later, in 2013 the same research group sequenced six more C. concisus strains using the “Illumina HiSeq” sequencing platform, and the de novo assembly tools “Velvet” and “SOAPdenovo”. The sequenced C. concisus strains were two reference strains ATCC 51561, ATCC 51562, and UNSWCS, UNSW1, UNSW2 and UNSW3. The metadata of C. concisus genome
sequences available in the database are presented in Table 1.2. A comparative genomic view of these eight *C. concisus* strains (Figure 1.11) is showing significant diversity among the strains (Deshpande *et al.*, 2013). The dendogram of the published *C. concisus* genomes is shown in Figure 1.12. In 2015, another reference *C. concisus* strain ATCC 33237\textsuperscript{T} isolated from gingival sulcus (the space between a tooth and the surrounding gingival tissue in the oral cavity) was fully sequenced and made available in NCBI (http://www.ncbi.nlm.nih.gov/nuccore/CP012541.1).

1.9.2 **Comparative genomic analyses of *C. concisus* and *C. jejuni**

A comparative analysis of eight *C. concisus* genomes with the nine available genomes of *C. jejuni* was performed by Deshpande *et al.* (2013). The core genome of the *C. concisus* strains was found to contain 1556 genes and the core genome of *C. jejuni* contained 1416 genes (Figure 1.13). The two sets of core genomes had 1033 shared genes between the two species, while 523 genes were specific to the *C. concisus* core and 383 specific to the *C. jejuni* core.
Table 1.2: *Campylobacter concisus* genome sequences available in NCBI database.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Disease</th>
<th>Genome size</th>
<th>% GC</th>
<th>Contigs</th>
<th>Scaffolds</th>
<th>Gene</th>
<th>Protein</th>
<th>Accession no</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(Mbp)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>C. concisus</em></td>
<td>Gastroenteritis</td>
<td>2.05 (Chr)</td>
<td>39.4</td>
<td>1</td>
<td>1</td>
<td>2062</td>
<td>1,983</td>
<td>CP000792</td>
</tr>
<tr>
<td>13826</td>
<td></td>
<td>0.03 (Plasmid)</td>
<td>31.6</td>
<td>1</td>
<td>1</td>
<td>29</td>
<td>29</td>
<td>CP000793</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.02 (Plasmid)</td>
<td>33.5</td>
<td>1</td>
<td>1</td>
<td>23</td>
<td>23</td>
<td>CP000794</td>
</tr>
<tr>
<td>ATCC 33237^T</td>
<td>Gingival sulcus</td>
<td>1.84</td>
<td>37.6</td>
<td>1</td>
<td>1</td>
<td>1,867</td>
<td>1,065</td>
<td>CP012541</td>
</tr>
<tr>
<td>ATCC 51561</td>
<td>Healthy control</td>
<td>1.99</td>
<td>39.6</td>
<td>69</td>
<td>69</td>
<td>2,025</td>
<td>1,900</td>
<td>ANNH00000000</td>
</tr>
<tr>
<td>ATCC 51562</td>
<td>Gastroenteritis</td>
<td>1.84</td>
<td>37.7</td>
<td>21</td>
<td>21</td>
<td>1,873</td>
<td>1,799</td>
<td>ANNI00000000</td>
</tr>
<tr>
<td>UNSWCD</td>
<td>Crohn’s disease</td>
<td>1.81</td>
<td>39.8</td>
<td>86</td>
<td>86</td>
<td>1,831</td>
<td>1,727</td>
<td>AENQ00000000</td>
</tr>
<tr>
<td>UNSW1</td>
<td>Gastroenteritis</td>
<td>1.94</td>
<td>39.5</td>
<td>72</td>
<td>72</td>
<td>1,963</td>
<td>1,858</td>
<td>ANNF00000000</td>
</tr>
<tr>
<td>UNSW2</td>
<td>Crohn’s disease</td>
<td>2.01</td>
<td>NA</td>
<td>98</td>
<td>25</td>
<td>NA</td>
<td>2,015</td>
<td>ANNJ00000000</td>
</tr>
<tr>
<td>UNSW3</td>
<td>Crohn’s disease</td>
<td>1.91</td>
<td>39.6</td>
<td>61</td>
<td>61</td>
<td>1,982</td>
<td>1,861</td>
<td>ANNE00000000</td>
</tr>
<tr>
<td>UNWCS</td>
<td>Gastroenteritis</td>
<td>2.11</td>
<td>39.2</td>
<td>207</td>
<td>53</td>
<td>2,286</td>
<td>2,085</td>
<td>ANNG00000000</td>
</tr>
</tbody>
</table>

NA: not available
The CGView tool used with *C. concisus* 13826 as the reference (outer 2 rings) highlights that the reference strain contains regions with high diversity when compared to the other seven *C. concisus* strains, adapted from Deshpande *et al.* (2013).
Figure 1.12: Dendogram of published *C. concisus* genomes based on genomic blast, adapted from (http://www.ncbi.nlm.nih.gov/genome/?term=Campylobacter+concisus).

Figure 1.13: The core genomes of *C. concisus* and *C. jejuni*, adapted from Deshpande *et al.* (2013).
This group has identified genes encoding for ferritin FtnA (involved in iron storage), and the Fe$^{3+}$ transport system FbpABC in *C. jejuni* for iron acquisition but not in *C. concisus* (Deshpande *et al.*, 2013). The Fe/Mn superoxide dismutase was present in *C. jejuni* while both the Cu/Zn and Fe/Mn superoxide dismutases were present in *C. concisus*. The flagellar P-rig protein PbpC encoding gene was only found in *C. concisus* which may associate with motility and adherence. The C4-dicarboxylate transport system DctPQM (sensing and differentiating between aerobic and anaerobic respiration) was found in the annotated core genome of *C. concisus* but was absent in the *C. jejuni* genomes which relates to a potential anaerobic growth characteristic of *C. concisus*. Major differences were found within respiration pathways of these two organisms. *C. concisus* genes of NorB (nitric oxide reductase) and NosZ (nitrous oxide reductase), were absent in *C. jejuni*. Yet, *C. jejuni* harbours genes of NrfA and NrfH (both are involved in nitrite respiration and NrfH functions to anchor NrfA into the membrane), while *C. concisus* contains genes encoding NrfC and NrfD (which have been linked to both nitrite and sulphite respiration).

There are similarities present as well, both organisms contain genes for NapG and NapH (two subunits of periplasmic nitrate reductase), *C. concisus* also contains a NapC/NirT cytochrome c family protein (membrane bound electron carrier), which is absent in *C. jejuni*. The authors also found that *C. jejuni* to have glutamate synthase GltD (glutamate synthase which converts 2-oxoglutarate to L-glutamate), while *C. concisus* contains GdhA (glutamate dehydrogenase which produces ammonia to convert 2-oxoglutarate to L-glutamate), which may indicate the respiring property of *C. concisus*. Some enzymes within the sulphur metabolism pathway were found in the genomes of the *C. concisus* strains but not in *C. jejuni* genomes. A sulphite reductase was present in *C. concisus* which
has the ability to convert sulphite to hydrogen sulphide which could be linked with the high level of H₂S production in UC patients (Deshpande et al., 2013). Furthermore, *C. concisus* core genome was molybdenum ion binding enriched and highly invasive *C. concisus* strains shared a molybdopterin synthase that was absent in the other strains (Deshpande et al., 2013).

### 1.9.3 The ‘pan and core’ genome of *C. concisus*

The pan-genome (or supra-genome) describes the full complement of genes in a particular species and the core genome is defined as a set of genes found in all sequenced genomes of a species. Analysis of pan and core genomes in eight *C. concisus* strains was determined including the reference sequence of *C. concisus* 13826 and other seven WGS (Deshpande et al., 2013). The core genome of *C. concisus* contained 1556 genes including 4.5% hypothetical proteins, but none of these core proteins were specific to only *C. concisus*. The pan genome was found to consist of 3254 genes, with many genes specific for only a sub-set of the strains, but not present in others. It was also found that the core genome contained a relatively large number of cytosolic proteins, metal-ion binding proteins and ATP binding proteins as these are house-keeping genes, essential for survival. On the other hand, the pan genome was enriched with defence responses, clearance of foreign intracellular DNA, DNA restriction-modification and DNA integration proteins. With these defence mechanisms the bacteria can protect itself from the invading foreign DNA by cleaving it with a restriction endonuclease. The self-DNA is protected by methylation of a specific nucleotide, which occurs immediately following replication, in the same target site of the restriction endonuclease. The bacterium mediates the insertion of foreign genetic material into the chromosome by the DNA integration proteins, or plasmids, and the incorporation of zot in the chromosome is an example of this mechanism.
In addition, the enrichment of the defence responses in the *C. concisus* pan genome indicated that this process is not conserved among all strains.

### 1.9.4 Plasmids in *C. concisus*

The number of plasmids present in *C. concisus* is variable. The reference strain *C. concisus* 13826 has two plasmids (pCCON31 and pCCON16) (Table 1.2), while the other reference strain ATCC 33237 does not contain any plasmid. The plasmids should possess an origin of replication (*ori*), generally made up of approximately four tandem repeats, flanked by an AT-rich region. In UNSWCD only one plasmid has been detected, however, no ‘*ori*’ was detected in any of its contigs. Twenty six genes were predicted in UNSWCD plasmid but none in pCCON31 and only 5 genes were conserved in pCCON16 (Deshpande *et al.*, 2011).

In the other six *C. concisus* WGS, synteny (conservation of blocks of order within two sets of chromosomes that are being compared with each other) and orthologous hit was compared with plasmids in *C. concisus* 13826 and UNSWCD. UNSW1 (22/26) and UNSW3 (24/26) have a higher level of conservation and synteny with the plasmid in UNSWCD. Interestingly, these three *C. concisus* strains were isolated from chronic disease patients. A lower level of conservation and synteny was observed when plasmid pCCON31 was compared with five strains (UNSW1, UNSW2, UNSW3, ATCC 51562 and ATCC 51562) with the exception of UNSWCS (24/33). In UNSWCS 12 of these 24 shared genes with pCCON31 were from one scaffold. Plasmid pCCON16 had a lower level of conservation and synteny with five strains (UNSW1, UNSW2, UNSW3, ATCC 51562 and UNSWCS) with the exception of ATCC 51561 (17/23) and these 17 shared genes that were found in two different scaffolds. It was suggested that the level of
pCCON16 conservation within the strains may correlate with their level of invasiveness of the strains (Deshpande et al., 2013).

### 1.9.5 Phylogenetic analysis based on ribosomal RNA (rRNA) genes

The 16S rRNA and 23S rRNA genes of the eight *C. concisus* strains were analysed to examine their evolutionary relationship and generate phylogenetic trees. The tree based on the 16S rRNA clustered the strains isolated from CD patients separately from the gastroenteritis isolates. The authors suggested that 16S rRNA was a suitable marker of genetic heterogeneity within the species, driven by evolution of different disease-causing *C. concisus* strains and the 23S rRNA tree showed two distinct clades, divided by genomospecies A or B (Deshpande et al., 2013).

### 1.10 Aims of the study

- To investigate the genotypic and phenotypic properties of biofilm formation of *Campylobacter concisus*.
- To study the influence of LuxS (S-ribosylhomocysteinase) on biofilm formation and other virulence characteristics.
- To sequence and perform comparative genomic analysis of the genomes of selected intestinal and oral *C. concisus* strains.
2.1 General Procedures

All chemicals and reagents used were of analytical laboratory reagent grade. Glassware was washed in Pyroneg detergent (Diverey Pty Ltd, Melbourne, Australia), rinsed twice in tap water, and then in deionised water before use. All solutions were prepared in deionised water delivered from a Millipore Milli-Q-water system (Liquipure, Melbourne, Australia). All media, reagents, pipette tips, microcentrifuge tubes and glassware were sterilised by autoclaving at standard conditions (121°C for 15 min), unless stated otherwise. All solutions were dispensed using Finnpipette (PathTech, Australia) for all volumes ranging from less than 0.5 µl to 5000 µl and were calibrated regularly as per manufacturer’s recommendations.

2.2 Bacteriological Methods

2.2.1 General materials

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetic acid, glacial</td>
<td>Sigma-Aldrich Pty., Ltd., Australia</td>
</tr>
<tr>
<td>Acetone</td>
<td>Sigma-Aldrich Pty., Ltd., Australia</td>
</tr>
<tr>
<td>Agar (Bacteriological Agar No. 1)</td>
<td>Oxoid Australia Pty. Ltd. Agarose</td>
</tr>
<tr>
<td>Agarose (DNA grade)</td>
<td>Bioline, Australia</td>
</tr>
<tr>
<td>Albumin, bovine serum</td>
<td>Sigma-Aldrich Pty., Ltd., Australia</td>
</tr>
<tr>
<td>Ammonium acetate</td>
<td>Ajax Chemicals, Australia</td>
</tr>
<tr>
<td>Ammonium chloride</td>
<td>BDH Chemicals, Australia</td>
</tr>
</tbody>
</table>

64
Ammonium hydroxide: BDH Chemicals, Australia

Ampicillin: CSL, Melbourne, Australia

Bacteriological tryptone: Oxoid Australia

Brain-Heart infusion broth: Oxoid, Australia

Brucella Broth: Oxoid, Australia

Calcium chloride (dihydrate, AR): BDH Chemicals, Australia

Cell scrapers (25 cm): Interpath Pty Ltd., Australia

Centricon Microconcentrator Starter Kit: Amicon, USA

Centrifuge tubes:

(i) 1.5 ml Eppendorf centrifuge tubes: Interpath, Australia

(ii) 10 ml centrifuge tubes: Interpath, Australia

(iii) 50 ml centrifuge tubes: Interpath, Australia

Columbia agar base: Cell BioSciences, Australia

Cover slips: Mediglass, Australia

Countess™ Automated Cell Counter: Invitrogen, Australia

Cryovials (1.8 ml): Nalgene Company, USA

Deoxynucleoside triphosphates (dNTPs): Bioline, Australia

Dextrose (D-glucose, AR): Ajax Chemicals Ltd, Australia

Dimethyl sulfoxide (DMSO): Sigma-Aldrich Pty., Ltd., Australia

DNA Ligase (T4): New England Biolabs, Australia

DNase I (bovine pancreas, grade I): New England Biolabs, Australia

DNA Polymerase: MyTaq DNA polymerase: Bioline, Australia

Dulbecco's Modified Eagle's Medium (DMEM): Thermofisher, Australia

Electroporation Cuvettes (0.2 cm): Molecular Bio Products, USA
## Chapter 2: General Materials and Methods

<table>
<thead>
<tr>
<th>Material/Reagent</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethanol</td>
<td>Merck Chemicals, Australia</td>
</tr>
<tr>
<td>Ethidium bromide</td>
<td>Sigma-Aldrich Pty., Ltd., Australia</td>
</tr>
<tr>
<td>EDTA (Ethylene diamine tetra acetic acid)</td>
<td>Merck Chemicals, Australia</td>
</tr>
<tr>
<td>Fetal Bovine Serum (FBS)</td>
<td>Invitrogen, Australia</td>
</tr>
<tr>
<td>Filters:</td>
<td></td>
</tr>
<tr>
<td>Syringe Filters (0.22µm, 0.45µm)</td>
<td>Gelman Sciences, USA</td>
</tr>
<tr>
<td>Ultrafiltration-unit filters (XM, YM)</td>
<td>Amicon, USA</td>
</tr>
<tr>
<td>Folin-Ciocalteau reagent</td>
<td>Ajax Chemicals Ltd, Australia</td>
</tr>
<tr>
<td>Formaldehyde</td>
<td>Merck Chemicals, Australia</td>
</tr>
<tr>
<td>Fluoro-Gel mounting medium</td>
<td>ProSciTech Pty Ltd, Australia</td>
</tr>
<tr>
<td>Gentamicin sulphate</td>
<td>Sigma-Aldrich Pty., Ltd, USA</td>
</tr>
<tr>
<td>Glucose</td>
<td>Merck Chemicals, Australia</td>
</tr>
<tr>
<td>Glycerol</td>
<td>Merck Chemicals, Australia</td>
</tr>
<tr>
<td>Horse Blood (Defibrinated)</td>
<td>Eqicell Products P/L., Australia</td>
</tr>
<tr>
<td>Hydrochloric acid (32 %)</td>
<td>Ajax Chemicals Ltd, Australia</td>
</tr>
<tr>
<td>Hydrogen peroxide (30 %)</td>
<td>BDH Chemicals, Australia</td>
</tr>
<tr>
<td>Immersion oil (Gurr)</td>
<td>BDH Chemicals, Australia</td>
</tr>
<tr>
<td>IPTG (Isopropylthio-β-D-galactosidase)</td>
<td>Bioline, Australia</td>
</tr>
<tr>
<td>Isolate II PCR and gel extraction kit</td>
<td>Bioline, Australia</td>
</tr>
<tr>
<td>Isolate II Plasmid extraction kit</td>
<td>Bioline, Australia</td>
</tr>
<tr>
<td>Isopropanol</td>
<td>Merck Chemicals, Australia</td>
</tr>
<tr>
<td>Kanamycin</td>
<td>CSL, Melbourne, Australia</td>
</tr>
<tr>
<td>Lambda DNA</td>
<td>New England Biolabs, Australia</td>
</tr>
<tr>
<td>MacConkey agar</td>
<td>Oxoid Australia</td>
</tr>
<tr>
<td>Magnesium chloride (Hexahydrate)</td>
<td>BDH Chemicals, Australia</td>
</tr>
<tr>
<td>Methanol</td>
<td>BDH Chemicals, Australia</td>
</tr>
</tbody>
</table>
Microaerobic gas mixture  
Microscope slides  
Microtitre plate (96 wells, flat bottom)  
Needle (sterile 18G, 19G, 21G)  
Nuclease-free Water  
Nutrient broth No. 2  
Paraffin  
PCR Master Mix  
Petri dish  
Phosphate buffer saline (PBS) tablets  
Potassium chloride  
Primers  
Proteinase K  
RNase A  
Skim milk  
Sodium acetate  
Sodium bicarbonate (7.5 %)  
Sodium chloride  
Sodium dodecyl sulfate (SDS)  
Di-Sodium hydrogen orthophosphate (anhydrous)  
Sodium hydroxide (pellets, AR)  
Sulphuric acid  
Syringe (1 ml, 5 ml, 10 ml, 20 ml, 50 ml)  
Tissue culture flask (25 cm², 75 cm²)  
Tissue culture Plate (6, 24 wells)  
TOPO TA cloning Kit  
Tris-base  

Linde Gas, Australia  
LOMB Scientific Co., Australia  
Corning, Australia  
Terumo Pty, Ltd., Australia  
Qiagen, Australia  
Oxoid Australia  
InterPath, Australia  
Bioline, Australia  
VWR international, Australia  
Oxoid limited, England  
BDH Chemicals, Australia  
Geneworks  
Sigma-Aldrich Pty., Ltd, Australia  
Sigma-Aldrich Pty., Ltd, Australia  
Bonlac Foods Ltd., Australia Skirrow  
BDH Chemicals, Australia  
Cytosystems Pty Ltd, Australia  
Merck Chemicals, Australia  
Crown scientific, Australia  
BDH Chemicals, Australia  
Merck Chemicals, Australia  
Merck Chemicals, Australia  
Terumo Pty, Ltd., Australia  
Nunc, Denmark  
Corning, Australia  
Invitrogen, Australia  
Merck Chemicals, Australia
Chapter 2: General Materials and Methods

Tris-HCl
Merck Chemicals, Australia

Triton-X-100
Sigma-Aldrich Pty., Ltd, USA

Trypsin – EDTA
Trace Biosciences, Australia

Tryptone
Oxoid, Australia Pty. Limited

Tween-20
BDH Chemicals, Australia

X-gal
Bioline, Australia

Yeast Extract
Oxoid, Australia

2.2.2 General equipment

<table>
<thead>
<tr>
<th>Equipment</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anaerobic jars</td>
<td>Oxoid Australia Pty. Ltd.</td>
</tr>
<tr>
<td>Balance:</td>
<td></td>
</tr>
<tr>
<td>(i) Analytical balance</td>
<td>Sartorius GMBH, Germany</td>
</tr>
<tr>
<td>(ii) Balance (0.1-500 g)</td>
<td>U-Lab, Australia</td>
</tr>
<tr>
<td>Cell counting chamber</td>
<td>Invitrogen, Australia</td>
</tr>
<tr>
<td>Centrifuge:</td>
<td></td>
</tr>
<tr>
<td>(i) Microcentrifuge (EBA12)</td>
<td>Zentrifugen, Germany</td>
</tr>
<tr>
<td>(ii) Bench top centrifuge (Centaur 2)</td>
<td>Graykon Scientific</td>
</tr>
<tr>
<td>(iii) High-speed centrifuge (L2-21 M/E)</td>
<td>Beckman, USA</td>
</tr>
<tr>
<td>(iv) Ultra-speed centrifuge (L8-80M)</td>
<td>Beckman, USA</td>
</tr>
<tr>
<td>Thermocycler</td>
<td>G-storm, UK</td>
</tr>
<tr>
<td>Electrophoresis Power Supply:</td>
<td></td>
</tr>
<tr>
<td>(i) PAC300</td>
<td>Bio-Rad Laboratories, USA</td>
</tr>
<tr>
<td>(ii) PPBasic</td>
<td>Bio-Rad Laboratories, USA</td>
</tr>
<tr>
<td>Electrophoresis Units (DNA):</td>
<td></td>
</tr>
<tr>
<td>(a) Mini gel (GNA-100)</td>
<td>Bio-rad Laboratories, USA</td>
</tr>
</tbody>
</table>
(b) Midi gel (wide mini-sub cell GT) Bio-rad Laboratories, USA

GELDOC system Bio-rad Laboratories, USA

Incubators

Tissue culture (5 % CO2) Forma Scientific, USA

Microscopes

(i) Light microscope Olympus Optical Co., Japan
(ii) Phase contrast microscope Nikon Kogaku KK, Japan
(iii) Confocal Microscope Nikon Eclipse Ti Inverted, Japan
(iv) Scanning electron microscope Quanta 200 SEM EM; FEI Co., Inc., OR

pH meter Radiometer, Denmark

Transilluminator (UV) BioRad, USA

Vortex mixer Ratek Instruments, Australia

2.2.3 General solutions

Agarose

Prepared as 1.0-1.5% (w/v) multipurpose agarose (Bioline).

Ethanol

Prepared as 70% (v/v) ethanol from absolute ethanol (96%), commercial grade (Merck).

Ethidium bromide (EtBr)

Prepared as 10 mg/ml EtBr solution in Milli-Q water.

Formaldehyde (2%)

Prepared by diluting formaldehyde solution (40%) (Merck, Australia) in Milli-Q water.

0.1M IPTG stock solution

Prepared by dissolving 1.2 g of isopropyl β-D-thiogalactopyranoside in 50 ml of sterile water.
6X loading buffer (for DNA gels)

Prepared as 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF and 40% (w/v) sucrose in Milli-Q water (Sambrook and Russell, 2001).

Lambda (λ) PstI DNA ladder

Prepared as 100 μg DNA (500 μg/ml, Promega), 100 U PstI (10 U/μl, Promega), 90 μl 10x restriction enzyme Buffer H (Promega), MQ water to 900 μl; incubated overnight at 37°C, then added 100 μl 10x gel loading dye; stored at -20°C.

Phosphate buffered saline (PBS)

Prepared by dissolving one PBS tablet in 100 ml Milli-Q water. The buffer was sterilised by autoclave.

RNase (DNase free)

Prepared as 20 mg/ml bovine pancreatic ribonuclease in sterile Milli-Q water, then heat treated at 100°C for 10 min. Aliquots of 100 μl each were kept frozen at -20°C.

50x TAE buffer (electrophoresis buffer for DNA gels)

Prepared by mixing 24.2% (w/v) Tris-base, 5.17% (v/v) glacial acetic acid and 1.86% (w/v) EDTA, then made up to 1 L in deionised water.

TE buffer

Prepared as 10 mM Tris base, 1 mM EDTA, pH 8.0 in Milli-Q water.

X-Gal Stock Solution (20mg/ml)

Prepared by dissolving 100 mg of 5-bromo-4-chloro-3- indolyl-β-D-galactoside (X-Gal) in 5 ml of N,N’-dimethylformamide. The solution was stored in the dark at -20°C.
2.2.4 Antibiotic stock solutions

Antibiotic solutions were prepared by dissolving the antibiotics in appropriate solvents to the desired concentration. These stock solutions were stored at -20°C after being filter sterilised. Antibiotic stock concentrations were as follows:

**Ampicillin**

Prepared as 100 mg/ml ampicillin (CSL) in sterile Milli-Q water; filter sterilized by passing through 0.2 μm filter (Whatmann syringe filter); stored at -20°C.

**Chloramphenicol**

Prepared as 30 mg/ml chloramphenicol (CSL) in 95% EtOH solution; filter sterilized by passing through 0.2 μm filter (Whatmann syringe filter); stored at -20°C.

**Gentamycin**

Gentamycin (Sigma Chemicals) was prepared as a stock of 10 mg/ml and filter sterilized by passing through 0.2 μm filter; stored in 1 ml aliquots at -20°C.

**Kanamycin**

Prepared as 50 mg/ml Kanamycin (Sigma Chemicals) in sterile Milli-Q water; sterilized by 0.2 μm filter (Whatmann syringe filter); stored at -20°C.

**Penicillin-Streptomycin**

Prepared as 5 mg/ml Penicillin-streptomycin (Thermo Trace) in DMEM medium used for tissue culture cell line maintenance.

2.2.5 Bacteriological media

All media were prepared according to the manufacture’s directions unless otherwise stated. Sterilization was performed at standard conditions, 121°C for 15 min. Media was poured into sterile Petri plates under aseptic conditions and left to set for 20
min in a laminar flow cabinet. Liquid media was used after autoclaving and cooling to room temperature.

### 2.2.5.1 Columbia blood agar (Horse blood agar)

Columbia agar base (4.1 % w/v) was dissolved in deionised water, autoclaved under standard conditions and cooled to 50°C–55°C before adding 10% (v/v) of defibrinated horse blood. The media was mixed and if required appropriate concentration of antibiotic was added.

### 2.2.5.2 Columbia broth

Columbia broth powder 3.5 % (w/v) dissolved in deionised water, autoclaved at standard conditions.

### 2.2.5.3 Brucella broth

Brucella broth powder (2 % w/v) was dissolved in deionised water and autoclaved at standard conditions, 1-2% yeast extract was added to enhance growth of *C. concisus*.

### 2.2.5.4 Brain heart infusion broth

Brain Heart infusion powder (3.7% w/v) was dissolved in distilled water and sterilized by autoclaving at 121°C for 15 min.

### 2.2.5.5 Brain heart infusion agar

Dehydrated Brain Heart Infusion broth (Oxoid) (3.7% w/v) and Bacteriological agar (1.5%) were dissolved in water and sterilised under standard conditions.
2.2.5.6 *Luria bertani agar (LA)*

Tryptone (1 % w/v), yeast extract (0.5% w/v), NaCl (0.5 % w/v) and bacteriological agar (1.5% w/v) were dissolved in deionised water and autoclaved at standard conditions.

2.2.5.7 *Luria bertani broth (LB)*

Tryptone (1 % w/v), yeast extract (0.5% w/v) and NaCl (0.5 % w/v) were dissolved in deionised water, dispensed into aliquots and autoclaved at standard conditions.

2.2.5.8 *SOC broth*

Tryptone (2 % w/v), yeast extract (1 % w/v), sodium chloride (10 mM), potassium chloride (2.5 mM), magnesium chloride (10 mM), magnesium sulphate (10 mM) were dissolved in water and autoclaved at standard conditions. Fresh filtered glucose (20 mM) was added prior to use.

2.2.5.9 *Muller hinton agar*

Muller Hinton Agar powder (Oxoid) (3.8% w/v) was dissolved in distilled water and autoclaved at standard conditions.

2.2.5.10 *Muller hinton broth*

Muller Hinton broth powder (Oxoid) (2.1% w/v) was dissolved in distilled water and autoclaved at standard conditions.

2.2.5.11 *Tryptone milk (Storage Media)*

Skim milk (10 % w/v) and tryptone (1% w/v) were dissolved in 10 mM Tris HCl (PH 7.5). The medium was autoclaved at 109°C for 30 min.
2.2.5.12 X-gal Plates

X-gal plates were prepared by spreading 40 µl of X-gal (20 mg/ml) and 4 µl of IPTG (0.1M) onto LB agar plates and dried before use. The plates were stored in the dark at 4°C.

2.2.6 Bacterial isolates and plasmids

Both clinical and oral isolates of *Campylobacter concisus* were used in this study. Clinical *C. concisus* strains were isolated from children suffering from gastroenteritis, at The Royal Children's Hospital, Melbourne, Australia (Ethics No. CA28021) and from IBD patients at Austin Health, Melbourne, Australia (Ethics No. H2012/04518). The oral *C. concisus* strains were isolated from volunteers (Ethics No. RMIT HERC No. 57/13). The list of the *C. concisus* and other bacterial strains are listed in Table 3.2. Plasmids used in this study are listed in Section 4.2.2.

2.2.7 Bacterial growth condition

*Campylobacter concisus* and *C. mucosalis* strains were grown on HBA for 48-96 h at 37°C under microaerophilic conditions, in a gas mixture consisting of 6% O₂, 8% CO₂, 6% H₂ and 80% balance of nitrogen in anaerobic jars without an anaerobic catalyst. *C. jejuni* and *C. coli* strains were grown on HBA plates in an atmosphere of 5% O₂, 10% CO₂ and 85% N₂ for 48 h at 37°C. Liquid cultures all *Campylobacter* spp. strains were grown in Brucella broth, supplemented with 2% yeast extract and incubated in jars with the same gas mixture as for plate cultures, at 37°C for 36-72 h. *Escherichia coli* strains were grown on nutrient agar (NA) for 18 h at 37°C. They were then inoculated into nutrient broth and grown for a further 18 h at 37°C, on an orbital shaker at 220 rpm. *E. coli* clones were cultured on Luria Bertani Agar (LA) with appropriate antibiotics while liquid cultures
Chapter 2: General Materials and Methods

were grown into Luria Bertani Broth (LB) (with appropriate antibiotic as needed), under the same previous conditions.

2.2.8 Storing bacterial strains

Bacterial strains and clones were harvested from lawn cultures with tryptone milk. Aliquots of cell suspension were placed into cryovials and stored at -80°C.

2.2.9 Measurement of bacterial cell concentrations

The cell density for bacterial cultures was estimated by absorbance readings (A_{600}) at 600nm. Absorbance reading was plotted against a growth curve for *Campylobacter* spp. and also confirmed by viable counts of bacterial dilutions on HBA plates. The growth medium was used as blank for absorbance readings.

2.3 Microscopy

2.3.1 Phase contrast microscopy

Bacterial suspensions were prepared in Brucella broth to a concentration of \( \sim 10^8 \) CFU/ml and dispensed into six well flat bottom culture plate (polystyrene) (Corning, Australia) and incubated under hydrogen enriched microaerophilic condition (6% O\(_2\), 10% CO\(_2\), 10% H\(_2\), and balanced N\(_2\)) for 5 d to observe the growth and development of biofilms attached to the polystyrene surface. Plates were incubated with a sterile cover slip (~ 0.5 cm x 0.5 cm) embedded within the well. Unbound planktonic cells in the wells were removed by a pipette, very gently washed twice with 0.9% saline (NaCl) to remove planktonic bacteria and 100 µl of 4% formaldehyde was dispensed to keep the biofilm moist. Cells attached to the inner surface of the glass cover slip were viewed by phase contrast microscopy using an Olympus CKX41 microscope with WHN110X-H/22
eyepiece and a 40X magnification LCACHN40XPH objective lens with a total of 400x magnification. All images were captured using a digital camera (Olympus Imaging Corp., Japan, model no. E-330 DC 7.4V) and stored as separate digital files for subsequent retrieval and analysis.

2.3.2 Confocal laser scanning microscopy (CLSM)

Bacterial suspensions prepared and dispensed into six well flat bottom culture plates as explained above. Plates were incubated with a sterile cover slip (~ 0.5 cm x 0.5 cm) embedded within the well. After incubation, the plate was rinsed three times with 0.9% saline (NaCl) to remove planktonic bacteria. The biofilm was stained using LIVE/DEAD BacLight bacterial viability kit (Invitrogen) (3.35 μM of Syto-9 and 20 μM of PI) at 25°C for 30 min in dark. The wells were washed twice very gently with normal saline (0.9% NaCl) to remove the unbound stain. The cover slip was removed very carefully and placed upside down on one drop of mounting oil (ProSciTech Pty Ltd, Australia) on a microscopic slide. The slide was examined immediately to observe the biofilm’s structure. To minimize artefacts associated with simultaneous dual wavelength excitation, all samples were sequentially scanned, frame-by-frame, first at 482 nm and then at 635 nm. The biofilms were then examined by a Nikon Eclipse Ti inverted microscope equipped with Nikon A1R Fast Laser Scanning Confocal system.

2.3.3 Scanning electron microscopy (SEM)

The biofilm morphology of the various C. concisus strains was studied using scanning electron microscopy. Biofilms for microscopy were cultured and incubated as described in Section 2.3.1. However, in this case, round shaped glass cover slips were used as the surface for biofilm formation. Once the biofilms had been developed, the cover slips
Chapter 2: General Materials and Methods

were removed very gently from the growth media and immediately fixed by karnovsky fixative, in 2% paraformaldehyde and 2.5% glutaraldehyde in 0.1 M sodium cacodylate buffer (pH 7.0) (Electron Microscopy Sciences) for at least 30 min at room temperature. After fixation, samples were washed with the same buffer 3 times and then dehydrated in series of ethanol solutions (50%, 70%, 90%, 95% and 2X 100%). They were subsequently air-dried in a ventilated covered container for at least 1 h and sputter-coated with a thin layer of gold in an Edwards S150 sputter coater. Next, the cover slip mounted on SEM stubs of a scanning electron microscope (Quanta 200 scanning electron microscope, EM; FEI Co., Inc., Hillsboro, OR) was used at high vacuum in the secondary electron imaging mode to observe the biofilms at 2,500x, 10,000x and 20,000x magnifications.

2.4 DNA Techniques

2.4.1 Genomic DNA extraction and purification

Genomic DNA was extracted using the Wizard Genomic DNA extraction kit (Promega, USA) according to the manufacturer’s instructions. Briefly, lawn cultures of *C. concisus* strains were grown under microaerophilic conditions for 48-72 h at 37°C. Cells were collected in a microcentrifuge tube in 1 ml of PBS, washed and centrifuged to get the pellet. Cells were lysed with the Nuclei Lysis solution provided with the kit and incubated for 5 min at 80°C. Then the cells were treated with RNase and incubated for 15-60 min at 37°C. The proteins were precipitated with Protein precipitation solution by incubation in ice for 5 min. After centrifugation, the supernatant was mixed with isopropanol to precipitate the genomic DNA. This DNA was cleaned by washing with 70% ethanol, air dried for 10-15 min and dissolved in rehydration solution. The purified genomic DNA was stored at -20°C for further use.
2.4.2 Plasmid DNA extraction and purification

Plasmid DNA was extracted using ISOLATE II plasmid mini kit (Bioline) according to the manufacturer’s instructions. Briefly, 1-5 ml of overnight bacterial culture was used for plasmid extraction. Cells were pelleted and resuspended in buffer P1 (RNAase added) and lysed in buffer P2. The reaction was stopped by adding neutralisation buffer P3 and centrifuged. The supernatant was loaded in the spin column washed with buffer PW1 (pre-warmed at 50°C) and PW2. Finally the plasmid DNA was eluted by the elution buffer and stored at -20°C.

2.4.3 Rapid boiled DNA extraction method

Bacterial suspensions in distilled water containing around $10^7$ CFU/ml were prepared as described earlier. One ml of this suspension was boiled for 5 min and then cooled on ice for 10 min to disrupt the cells. The bacterial debris was removed by centrifugation at 10,000 x g for 5 min. Ten µl of the DNA containing supernatant were used to provide the template DNA for PCR reactions.

2.4.4 Quantification of DNA concentration

The concentration of the extracted DNA (chromosomal DNA and plasmid DNA) was measured by appropriate methods. Concentrations of plasmid DNA were estimated by comparison of sample DNA against the lambda PstI digested molecular weight marker (50 µg/ml). For chromosomal DNA, concentrations were estimated by measuring the optical density at wavelengths 260 nm and 280 nm. Genomic DNA concentrations were measured using a Nanodrop Lite Spectrophotometer (Thermo Scientific). The purity of the sample was determined by the ratio of ODs obtained at 260 nm and 280 nm (>1.8 for clean DNA).
2.4.5 Polymerase Chain Reaction (PCR)

PCR was performed by a G-storm Thermal Cycler Machine in sterile PCR tubes. Colony PCR was conducted by using fresh isolated colonies which were taken up by sterile tips. The colonies were immersed into 50 µl PCR reaction mix and mixed by pipetting up and down few times. The remaining bacterial cells in the colony were subcultured on an agar plate with an appropriate antibiotic. PCR Master Kit supplied by Bioline (Australia) was used to set up the reaction. PCR was performed using the general protocol described in the instruction manual. Annealing temperature of the reaction was determined by calculating the Tm of the primers. The details of the PCR cycling conditions are stated in each section.

2.4.5.1 Primers design

Primers were designed using the Clone Manager suite of analysis tools (version 7.11) (Sci Ed Central website). Clone Manager was also used to confirm that the genes and corresponding designed primers were correctly cloned into the vector, both in frames with the start codon of the vector and in the correct orientation for proper expression of the selected gene. All primer sequences were selected from the genome sequence of *C. concisus* 13826 or stated otherwise. Restriction enzyme (RE) sites were also included in the primer design to facilitate the cloning of amplicons into the vectors and %GC contents were between 40-60% with a Tm in the range of 50-80°C.
2.4.5.2 General PCR protocol

All PCR were performed using the following reagents (if not stated differently):

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume (µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA template (8 ng/µl)</td>
<td>2.5</td>
</tr>
<tr>
<td>Primer forward (10 pmole/µl)</td>
<td>0.5</td>
</tr>
<tr>
<td>Primer reverse (10 pmole/µl)</td>
<td>0.5</td>
</tr>
<tr>
<td>My Taq reaction buffer</td>
<td>5</td>
</tr>
<tr>
<td>My Taq polymerase</td>
<td>0.125</td>
</tr>
<tr>
<td>Milli Q H₂O</td>
<td>16.4</td>
</tr>
<tr>
<td>Total volume</td>
<td>25.0</td>
</tr>
</tbody>
</table>

A negative control containing all previous reagents except the DNA template, which was replaced with an equal volume of sterile distilled water, was used in all PCR amplifications to ensure the PCR mix purity. All PCR were repeated 3 times.

2.4.5.3 Agarose gel electrophoresis

Agarose gel electrophoresis was used to separate DNA fragments by length and to estimate the size of DNA. Agarose gels were prepared using 1.0-2.0% of agarose powder in 1 x TAE buffer and dissolved by boiling in a microwave and poured into a casting tray after cooling to 50°C. An electric field was created by applying a current of 40 – 100 V for 1 – 3 h and the DNA fragments were separated according to size. The size of the DNA was estimated by electrophoresing each sample against a lambda DNA (λ DNA) marker prepared by digestion with PstI restriction enzyme (Appendix I). The gel tanks used were either mini or midi with 1x TAE buffer as the wells needed. The gels were stained in an
Chapter 2: General Materials and Methods

EtBr bath for 10 min and destained under running tap water for 30 min. The Gel Doc Imaging system (Bio-Rad) was used to visualize DNA fragments under UV illumination.

2.4.5.4 DNA purification from agarose gel and PCR product

PCR product and DNA bands from the gel were purified using ISOLATE II PCR and Gel extraction kit (Bioline) according to the manufacturer’s instructions.

2.4.5.5 DNA preparation for sequencing

The purified PCR product or the DNA band purified from gel or plasmid DNA was quantitated in a Nanodrop Lite Spectrophotometer (Thermo Scientific). Purified product (6-75 ng, according to the size of the PCR product; 600-1500 ng for double stranded plasmid) was mixed with either forward or reverse primers (9.6 pmole/µl) and sent to Australian Genome Research Facility Ltd (AGRF) for sequencing.

2.4.6 DNA Manipulation

2.4.6.1 Restriction digestion

All digestion reactions were carried out with the following reagents:

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>X µl</td>
</tr>
<tr>
<td>Restriction Enzyme</td>
<td>0.5 µl (10 units/µl)</td>
</tr>
<tr>
<td>10x buffer</td>
<td>2 µl</td>
</tr>
<tr>
<td>BSA</td>
<td>1 µl 1 (2mg/ml)</td>
</tr>
<tr>
<td>Water</td>
<td>X µl</td>
</tr>
<tr>
<td><strong>Total volume</strong></td>
<td><strong>20 µl</strong></td>
</tr>
</tbody>
</table>

All digestion reactions were carried out at 37°C for 2 - 4 h and then stored at -20°C.
2.4.6.2 DNA ligation

Ligation reactions were performed either incubating at 16°C overnight or room temperature for 10 min, using the following reagents:

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>10X T4 DNA Ligase Buffer</td>
<td>2 µl</td>
</tr>
<tr>
<td>Vector DNA</td>
<td>(x µl) 50 ng</td>
</tr>
<tr>
<td>Insert DNA</td>
<td>(x µl) 37.5 ng</td>
</tr>
<tr>
<td>Nuclease-free water</td>
<td>to 20 µl</td>
</tr>
<tr>
<td>T4 DNA Ligase</td>
<td>1 µl</td>
</tr>
<tr>
<td>Total volume</td>
<td>20 µl</td>
</tr>
</tbody>
</table>

The amount of plasmid DNA and insert DNA added to the ligation reaction was determined by the following equation:

\[
\text{ng of DNA insert} = \frac{\text{ng of vector} \times \text{size of insert (kb)}}{\text{Size of vector (kb)}} \times \text{molar ratio of insert to vector}
\]

2.4.7 Preparation of competent \textit{E. coli} DH5\(\alpha\) cells

2.4.7.1 Preparation of electro competent cells (\textit{E. coli} DH5\(\alpha\))

A single colony of \textit{E. coli} DH5\(\alpha\) was inoculated into 5 ml of LB broth and grown overnight at 37°C on a shaker. Two ml of the overnight culture was transferred to 200 ml of pre-warmed LB broth and the cells were grown at 37°C with vigorous shaking to an OD 600 of 0.6 (~2-3 h). The cells were then chilled on ice for 30 min to slow down the cell growth. The cells were harvested by centrifugation at 4700 g for 15 min at 4°C and the cell pellet was resuspended in 200 ml chilled sterile Milli-Q water. The supernatant was completely drained off after centrifugation and then resuspended in 100 ml of 10% glycerol in chilled sterile water and centrifuged as above. The cells were pelleted and
resuspended again in a final volume of 2 ml of ice cold 10% glycerol. Centrifugation was repeated and the pellet was resuspended in 250 μl of 10% glycerol and aliquoted out in 50 μl volumes and stored immediately at -80°C.

2.4.7.2 Preparation of chemical competent cells (E. coli DH5α)

Competent E. coli DH5α cells were prepared using a modification of the method described by Cohen et al. (1972). All solutions and centrifugation steps in this procedure were performed at 4°C. A single colony of E. coli DH5α from an agar plate was inoculated into 10 ml of LB broth and incubated on a shaker at 37°C for 16 h. One ml of this overnight culture was used to inoculate 30 ml of LB broth. The culture was grown to exponential phase by shaking vigorously at 37°C for 3 h, and then chilled on ice for 30 min. The broth culture was transferred into 15 ml centrifuge tubes and cells were collected by centrifugation at 4000x g for 5 min. The supernatant was decanted and the cells were gently resuspended in 5 ml of ice-cold MgCl₂ (100 mM). The cells were re-pelleted at 4000x g for 5 min and resuspended in 5 ml of 75 mM CaCl₂ per tube and incubated on ice for 30 min, then centrifuged as above. Competent E. coli DH5α cells were resuspended in 1 ml of 75 mM CaCl₂ and divided into aliquots of 100 μl each. Competent cells were either used immediately or frozen at -80°C for future.

2.4.7.3 Electrotransformation (Electroporation)

The method employed was based on the Biorad Gene Pulser apparatus instruction manual. Briefly, 1-2 μg of plasmid DNA was added carefully to a 50 μl vial of electrocompetent E. coli DH5α cells and stored on ice for 1 minute before transferring to a cuvette. The Gene Pulser setting used to deliver the DNA into the cell was set at 200 Ω, 25 μF capacitance, and 2.0-2.5 kV. The ligated product with plasmid DNA was transferred to
the bottom of an ice-cold sterile cuvette (with a 2 mm gap). The cuvette was placed in the apparatus using an ice-cold sliding device, and pulsed once. One ml of SOC medium or LB broth was immediately added to the cuvette resuspending the cells. The electroporated sample was transferred to a 1.5 ml polypropylene tube, and incubated for 1 h at 37°C with shaking. Hundred µl of suspension was plated out onto a pre-warmed LB agar plate containing selective antibiotics. The remaining 900 µl was centrifuged at 1200x g (RT), and the supernatant discarded. The cells were resuspended in 100 µl broth medium, and the entire 100 µl “concentrated transformants” was plated out onto another LB agar plate containing antibiotics. The plates were incubated at 37°C overnight, and checked for the transformants and stored at 4°C. Transformation efficiency was tested by adding 1 µl of undigested “empty” vector (miniprep sample) to a 50 µl vial of electrocompetent E. coli DH5α cells, and the cells were electro-transformed as described above.

2.4.7.4 Chemical transformation

Chemically competent E. coli DH5α were used to perform chemical transformation. Two µl of “ligated” product was added into one vial (50 µl) of chemically competent cells, mixed gently and incubated for 30 min on ice. The cells were exposed to a heat shock for 30 sec at 42°C (water bath) and immediately transferred to an ice bath. SOC medium (900 µl) was added into the tubes and incubated at 37°C for 1 h on a shaker. The transformants were plated out on to pre-warmed LBA plates with selective antibiotics. Transformants were visualised by blue/white colony screening.

2.4.8 Preparation of C. concisus competent cells

C. concisus competent cells were prepared by washing with ice cold filter sterilised sucrose glycerol solution. Sucrose-Glycerol solution was prepared by adding 9% sucrose
and 15% glycerol in Milli-Q water and filter sterilised by passing through a 0.2µm membrane filter. Bacterial cells were collected by centrifugation at 3,300x g for 5 min at 4°C. The pelleted cells were resuspended in 0.5 ml of cold sucrose glycerol solution and aliquoted in 40 µl volumes into microcentrifuge tubes. The competent cells were stored at -80°C.

2.4.8.1 Electrotransformation of plasmid DNA into competent C. concisus

The constructed plasmid DNA (0.1-1.0 µg) was mixed with C. concisus competent cells and kept on ice for 1 minute. The mixture was transferred into a chilled electroporation cuvette of 0.2 µm width and electroporated at 25 µF, 2.5 kV and 200 Ω using Gene Pulse Controller (Bio-Rad). The cuvette was immediately filled with 1 ml of SOC medium and transferred onto a HBA plate without antibiotic. The plates were incubated at 37°C under microaerophilic conditions for overnight. The cells were harvested in Heart Infusion Broth and grown on HBA plate supplemented appropriate antibiotics. The plates were incubated at 37°C for 2 days under microaerophilic conditions.

2.4.8.2 Natural transformation of C. concisus

Campylobacter transformation was performed either on an agar surface or in a biphasic system following the method described by Wang and Taylor (1990). C. concisus cells were grown on HBA plates at 37°C for 16 h. For transformation on MH agar surface, fresh recipient cells were spread on MH agar at about 5 x 10^7 cells per plate and incubated for 6 h. Aliquots of DNA (plasmid DNA or linear DNA) were spotted directly onto the inoculated agar without additional mixing or spreading, and incubation was continued for 5 h. For transformation in a biphasic system, the cells were resuspended in 1
ml of BHIB. Two hundred μl volumes of cells were over layered into a microcentrifuge tube with 1 ml of BHI agar. The lid was closed and pierced. The microcentrifuge tube with media was incubated under microaerophilic conditions for 2-6 h at 37ºC. Ten μg of plasmid DNA were added on to the bacterial suspension and incubated at 37ºC for another 3-5 h. The entire volume of cells and the DNA were spreaded on to HBA plates with appropriate antibiotics. The plate was incubated under microaerophilic conditions at 37ºC for 48-72 h. Colonies were confirmed by colony PCR and restriction enzyme digestion.

2.4.9 Genetic complementation

2.4.9.1 Construction of a shuttle vector

To complement the luxS-mutation, a second copy of luxS was inserted (under the control of its native promoter) into a second chloramphenicol resistant plasmid, pBC SKII + (Appendix IV). *E. coli*-C. concisus shuttle vector a was constructed by inserting a small plasmid (3316 bp) present in *C. concisus* RCH 26 carrying mob, repC, and repE and condensin complex subunit 2 genes. It was amplified with inverse primers inserting BamHI and PstI restriction enzymes sites and cloned into pBC SKII + containing luxS.

2.4.9.2 Transformation of the shuttle vector into a *C. concisus* mutant

The new constructed shuttle vector was transformed in the *C. concisus* mutant strain by natural transformation or electrotransformation as described in Section 2.3.8.1 and 2.3.8.2.

2.5 Tissue Culture Methods

All tissue culture related work was carried out in a biological safety cabinet class II to minimise risk of contamination. Glassware used for storage of media and solutions was
washed with Pyroneg detergent, rinsed twice in dH$_2$O and then soaked overnight in dH$_2$O. All glassware was autoclaved before being used to make and store tissue culture media. Solutions and dH$_2$O (for tissue culture media) were made in sterilised bottles before being re-autoclaved. A human intestinal epithelial cell line (INT 407) was cultured in standard DMEM media and standard PBS was used when working with INT 407 cell line.

2.5.1 Tissue culture Media and Solutions

2.5.1.1 Dulbecco's Modified Eagle Medium (DMEM)

To 372 ml of deionised distilled water (double sterilised), 100 ml of 5x DMEM, 13.5 ml of 7.5% Sodium bicarbonate, 10 ml of HEPES buffer and 4.5 ml of L-glutamine were added by direct filter sterilisation and stored at 4°C. Fresh aliquots of DMEM supplemented with 10% v/v newborn calf serum (NCS) were used in culture.

2.5.1.2 Trypsin-EDTA solution

Trypsin-EDTA stock was mixed with 1xPBS at a ratio of 1:1 and stored at -20°C. This solution was used to dissociate INT 407 cells from the flask.

2.5.1.3 Gentamicin solution

Gentamycin sulphate (400 μg/ml) in DMEM

2.5.1.4 Triton X-100 (detergent) 0.25%

The detergent Triton X-100 was diluted in sterile dH$_2$O to a concentration of 0.25% v/v. Solution was incubated at 37°C for 30 min, filter-sterilised and stored at room temperature.
2.5.2 Tissue culture techniques

2.5.2.1 Tissue culture growth conditions

Frozen tissue culture cells were resuscitated from liquid nitrogen, thawed and grown in their respective media in a 37°C CO₂ incubator. Cells were grown until 80-95% confluency was reached before being split and transferred to a new flask or used in an assay. Tissue culture cells were only passaged between 5-10 times before a new stock of cells was resuscitated.

2.5.2.2 Splitting, recovery and banking of cells

Old media was removed from the flask and the cells washed twice with 1xPBS. To dissociate INT 407 cells, an appropriate amount of trypsin-EDTA solution was added and the flask was incubated at 37°C for 10-15 min. The dissociated cells were washed twice in the appropriate media and collected each time by centrifugation at 1000 x g for 5 min. Cells were resuspended in 1 ml of DMEM/FBS and an appropriate amount was used to seed a new tissue culture flask.

Cells were banked down for storage as outlined above but resuspended in media supplemented with 10% DMSO, which acts as a cryoprotectant. The cell suspensions were placed into cryovials and stored at -70°C for 2-3 days before being transferred to liquid nitrogen for long term storage.

2.5.2.3 Seeding of tissue culture trays

Tissue culture cells were prepared according to the protocol for splitting the cells, as outlined above. A 10-fold and 100-fold dilution of cells was made with PBS (to a total volume of 100 μl). A volume of 92 μl of each dilution was mixed with 8 μl of 0.25%
trypan blue and incubated at room temperature for 5 min to allow dead cells to be stained. Each dilution was then transferred (10 μl) to a grid on a cell counting chamber and covered with a cover slip. Viable cells were counted (blue cells were excluded) and the number of cells required to seed each well was calculated.

The required amount of cells was mixed with the appropriate DMEM/FBS and 500 μl of this suspension was dispensed into each well in a 24-well tissue culture tray. This was then incubated at 37°C in 5% CO₂ for 16-18 h before being used in an assay.

2.5.2.4 Invasion/Adhesion assays

Adhesion and invasion assays of *C. concisus* strains in INT 407 cells were carried out based on the method previously described by Elsinghorst (1994) and modified by Jason Chang, RMIT University (Chang, 2002). Briefly, cells were cultured, harvested and counted. Cell line was seeded at a concentration of 2x10⁵ cells per well and incubated overnight. Before commencing the assay, media was aspirated from each well and the cells washed 3 times with PBS. To each well, 100 μl of ~ 10⁸ bacterial cells (MOI ~ 50-100 bacteria per tissue culture cell) along with 400 μl of DMEM/FBS was added and the culture trays were incubated at 37°C in 5% CO₂. Each assay was performed in triplicates and repeated three times on different days.

For the adhesion assay, tissue culture trays were incubated for 6 h as described above. Wells were then washed 3 times with PBS. Tissue culture cells were lysed by the addition of 100 μl of 0.25% Triton X-100 into each well and incubation at 37°C for 15 min. After pipetting 800 μl of PBS into each well, 100 μl was subjected to a serial 10-fold dilution before being plated out on HBA plates. Plates were incubated 48-72 h at 37°C in
presence of H$_2$ and bacteria enumerated. For the invasion assay, tissue culture trays were incubated for 6 h. Contents of each well was washed 3 times with PBS before 500 μl of 400 μg/ml of gentamicin solution was added. Culture trays were incubated at 37°C in 5% CO$_2$ for 1 h to kill any extracellular bacteria. Wells were again washed thrice with PBS and intracellular bacteria released with Triton X-100 as mentioned above. Dilutions were plated and counted as outlined above.

### 2.6 Whole Genome Sequencing (WGS) of *C. concisus* strains

Whole genome sequencing was performed for *C. concisus* strains by Ion Proton platform and Illumina Miseq platform at RMIT University.

#### 2.6.1 Preparation of the genomic DNA from *C. concisus*

*C. concisus* genomic DNA was extracted using Wizard Genomic DNA extraction kit (Promega, USA) according to the manufacturer's instructions as described in Section 2.3.1. The DNA sample was visualised using agarose gel electrophoresis (Section 2.3.5.3). The quality and quantity of the DNA sample was evaluated on the NanoDrop Lite spectrophotometer (Thermo Scientific, USA).

#### 2.6.1.1 Quantification of DNA concentration by Qubit

The concentration of DNA was measured using Qubit (Invitrogen by Life Technologies, Singapore) according to the manufacturer's instructions. Briefly, the Qubit® working solution was prepared by diluting the Qubit® dsDNA HS Reagent 1:200 in Qubit dsDNA HS Buffer. Qubit working solution was added with the DNA sample ranged from 1–20 μl to make the final volume up to 200 μl and mixed by vortexing 2–3 sec. The tubes were incubated at room temperature for 2 min and the concentration was measured in Qubit 3.0 Fluorometer (Life Technologies, Singapore).
2.6.2 WGS in Ion Proton Platform

2.6.2.1 Library preparation for Ion Proton sequencing

The genomic library preparations from *C. concisus* strains were performed using Ion Xpress™ Plus gDNA Fragment Library Preparation kit (Life Technologies, USA) according the manufacturer’s instructions. The workflow diagram of the library preparation procedures is presented in Figure 2.1 (adapted from Ion Torrent, 2013).

Based on the Qubit reading, 50-100 ng of genomic DNA (gDNA) was used for the library preparation. The gDNA was then fragmented using Ion Shear™ Plus Reagents (Life Technologies, USA). The fragmentation reaction time was optimised for the gDNA of *C. concisus* strains in order to produce fragments with a median size of 100-200 bp. Every library preparation step was followed by purification and cleaning of the DNA using Agencourt® AMPure® XP Reagent (Beckman Coulter, Inc., USA) according to the manufacturer’s instructions. The DNA fragments were then ligated to Ion-compatible adapters (Figure 2.1).

This was followed by nick-repair to complete the linkage between the adapters and DNA inserts. In the case of barcoded libraries, adapters from the Ion Xpress™ Barcode Adapter kits (Life Technologies, USA) were used. The adapter-ligated and nick-translated DNA was purified and cleaned before being size selected. The adapter-ligated library was size selected to have the optimum length according to the targeted read length of ~220 bp. The size selection was performed using E-gel® agarose gel (size select 2%, Invitrogen, Israel), E-gel® safe imager and real-time transilluminator (Invitrogen by Life Technologies, Israel). The final step was library amplification followed by purification and final cleaning of the library preparations.
Figure 2.1: Overview of the genomic DNA library preparation using the Ion Xpress™ Plus gDNA Fragment Library Preparation kit. ‘A’ and ‘P1’, Ion-compatible adaptors; ‘BC’, barcode (Adapted from Ion Xpress™ Plus Fragment Library Preparation User Guide).
2.6.2.2 Library preparation analysis using Bioanalyzer

The library preparations were then evaluated for the quality and quantity using a Bioanalyzer (Agilent® 2100, Germany) before use in template preparation steps. The High Sensitivity DNA Analysis Kit (Agilent Technologies, Germany) was used for the sample and chip preparations according to the manufacturer’s instructions. The Bioanalyzer (Agilent® 2100, Agilent Technologies, Germany) uses a chip-based nucleic acid separation system performing capillary electrophoresis on micro-fabricated channels. It performs detection and online data evaluation in an automated manner (Agilent Technologies, 2013). The data obtained from the Bioanalyzer were used to determine the molar concentration of the DNA samples and to evaluate the size distribution of the libraries. The DNA molar concentrations were then used to calculate the appropriate dilution of the library required for template preparation.

2.6.2.3 Template preparation

The sequencing template preparation for *C. concisus* strains was performed using the Ion OneTouch™ 2 system and Ion PI™ Template OT2 200 Kit v3 (Life Technologies, USA) according to the manufacturer’s instructions. The Ion OneTouch™ 2 System included the Ion OneTouch™ 2 instrument (Life Technologies, USA) and the Ion OneTouch™ ES instrument (Life Technologies, USA). The final library preparation with concentration of 8 pM was used for template preparation of *C. concisus* strains. Briefly, the template-positive ISPs, containing clonally amplified DNA, were prepared by emulsion-PCR using the Ion PI™ Template OT2 200 Kit v3 (Life Technologies, USA) with the Ion OneTouch™ 2 Instrument. Then the template-positive ISPs were enriched using the Ion OneTouch™ ES. The enriched ISPs were then collected and washed before being used in the sequencing reaction.
2.6.2.4 Sequencing run

The sequencing reaction was carried out using the Ion PI™ Sequencing 200 Kit v3 on the Ion Proton™ Semiconductor sequencer (Life Technologies, USA) according to the manufacturer’s instructions. In summary, a planned run was created using Torrent Suite™ software connected to the Ion Proton™ Sequencer. This specified the number of runs, kit types, barcode used, the run type (gDNA sequencing) and the reference sequences. The Ion Proton™ sequencer was cleaned using chlorite cleaning and water cleaning according to the instructions followed by the sequencer initialisation. In a separate experiment, the Ion PI™ Chip (v2) (Life Technologies, USA) was prepared for loading by repetitive washing with nuclease-free water (Invitrogen by Life Technologies, USA), isopropanol (Sigma-Aldrich, USA), Ion PI™ Chip preparation solution (Life Technologies, USA) and NaOH (0.1 M, freshly prepared). This was followed by the Ion PI™ Chip calibration according to the manufacturer’s instructions.

The template-positive Ion PI™ ISPs was prepared for sequencing by addition of the Ion PI™ Control Ion Sphere™ particles (Life Technologies, USA) to the enriched ISPs. This was followed by annealing the Ion PI™ Sequencing Primer (Life Technologies, USA) to the enriched ISPs. Then the chip was loaded with 55 μl of the template-positive ISPs with 10 min centrifugation using Ion Chip™ Minifuge (Ion Torrent, Life Technologies, Korea). The chip containing the sample was then flooded with annealing buffer (Life Technologies, USA) and the prepared foaming Solution (10% Triton® X-100, Life Technologies, USA). Lastly, the chip was flushed with the flushing solution (Life Technologies, USA) and the Ion PI™ Sequencing Polymerase (v3, Life Technologies, USA) was loaded on the chip. The sequencing run was then started.
2.6.3 WGS in Illumina Platform

2.6.3.1 DNA sample preparation

The indexed paired-end libraries from genomic DNA (gDNA) of *C. concisus* strains were performed using Illumina Nextera® DNA Sample Preparation Kit (Illumina, California, U.S.A.) for subsequent cluster generation and DNA sequencing. This protocol was performed to fragment and add adapter sequences onto template DNA with a single tube Nextera reaction (tagmentation) to generate paired end sequencing libraries. The workflow diagram of the library preparation procedures is presented in Figure 2.2 (Nextera® DNA Sample Preparation Guide, 2012).

Briefly, genomic DNA was tagmented (tagged and fragmented) by the Nextera transposome (Illumina, USA). The Nextera transposome simultaneously fragments the genomic DNA and adds adapter sequences to the ends. In this procedure a total of 50 ng gDNA was added with Tagment DNA Buffer in a 96-well Nextera Enrichment Tagmentation Plate, centrifuged and kept at 55°C for 5 min in a thermocycler, then continued to clean-up of tagmented DNA. In the next step, the tagmented DNA was purified from the Nextera transposome. Zymo DNA binding buffer (180 µl) was taken in a new Nextera Sample Plate 2 (NSP2). Fifty µl of tagmented DNA was transferred to NSP2. The mixture was transferred to Zymo-Spin™ I-96 Plate (Illumina, USA). The plate was centrifuged for 2 min and flow-through was discarded. The plate was washed twice. Then resuspended in buffer for 2 min and centrifuged to get the supernatant. Later the purified tagmented DNA was amplified via a limited-cycle PCR Program. The PCR step also adds index 1 (i7) and index 2 (i5) and sequencing, as well as common adapters (P5 and P7) required for cluster generation and sequencing. This step uses AMPure XP beads to purify
Figure 2.2: Illumina gDNA fragment library preparation overview, adapted from Nextera DNA Sample Preparation Guide, (Illumina, October 2012).
the library DNA, and provides a size selection step that removes very short library fragments from the population.

2.6.3.2 Library preparation analysis using Bioanalyzer

The library preparations were then evaluated for the quality and quantity using a Bioanalyzer (Agilent® 2100, Germany) as described in Section 2.6.2.2. Typical libraries should show a broad size distribution from ~250 bp to 1000 bp.

2.6.3.3 Reagent preparation

One single-use MiSeq® Reagent Kit v3 (Illimina, USA) was used to perform the run on the Illumina MiSeq. It contains the MiSeq reagent cartridge, hybridization Buffer, MiSeq Flow Cell and incorporation Buffer. MiSeq reagent cartridge is a single-use consumable consisting of foil-sealed reservoirs pre-filled with clustering and sequencing reagents sufficient for sequencing one flow cell. Hybridization Buffer was used to dilute denatured libraries before loading libraries onto the reagent cartridge for sequencing. Once the reagent cartridge is prepared following the instructions, it was preceded to sequence run.

2.6.3.4 Sequencing run

The prepared libraries were denatured and diluted for cluster generation and sequencing. Libraries were loaded onto the reagent cartridge in the reservoir labelled Load Samples. Using the MiSeq Control Software interface (Illumina, USA), following the run setup steps to load the flow cell and reagents, the run was started.
2.6.4 Sequencing data analysis and quality control

The raw sequencing data were downloaded on the Torrent Browser (Life Technologies). The initial quality control and analysis was performed using Torrent Suite™ software. A Torrent Browser run report was created for the sequencing run.

2.6.5 De novo genome assembly

The raw data generated from the Ion Proton and Illumina Miseq were extracted and de novo assembly was performed for each genome using Mira 4 (Chevreux et al., 2004), CLC Genomics Workbench 8.0.3 (https://www.qiagenbioinformatics.com/), and A5-miseq (Coil et al., 2015). Detail procedures of these assemblies are described in Chapter 5.

2.6.6 Annotation and gene prediction

The RAST web application server (Aziz et al., 2008) was used for gene predictions using the Glimmer program. Comparative genomics modules available in RAST were used for gene based comparisons between reference sequences C. concisus 13826 and ATCC 33237T and four sequences from this study. Genome scale alignments were performed using the Mauve alignment tool (Darling et al., 2004) and CGView (Stothard and Wishart, 2005).

2.6.7 Plasmid identification and bioinformatics analysis

The web-based tool ‘Tandem Repeat Finder’ (Benson, 1999) was used to check the presence of origins of replication (ori) within contigs less than approximately 40 kb size.
Chapter Three

Investigation of *Campylobacter concisus* for biofilm properties

3.1 Introduction

*C. concisus* is an oral organism and emerging gastroenteritis pathogen. It has been associated with acute intestinal disease (Engberg *et al.*, 2000; Lastovica and le Roux, 2000) as well chronic intestinal disease such as Crohn’s disease (CD) in children (Zhang *et al.*, 2009) and ulcerative colitis (UC) in adults (Mahendran *et al.*, 2011). On the other hand, it has been isolated from both healthy and diseased periodontal sites, though healthy sites have been shown to harbour higher levels of *C. concisus* (Macuch and Tanner, 2000). It is also associated with gingivitis (Socransky *et al.*, 1998), periodontitis (Kamma *et al.*, 1994), saliva from Inflammatory bowel diseases (IBD) patients (Zhang *et al.*, 2010). Despite its fastidious nature and sensitivity to environmental oxygen, *C. concisus* can survive in the oral cavity and the survival strategy is possibly formation biofilms at these sites. Indeed *C. concisus* has been shown to be capable of producing biofilms on different surfaces such as glass, stainless steel, and polystyrene (Gunther and Chen, 2009). More recently, it was reported that *C. concisus* isolates from patients with a range of intestinal diseases are able to produce biofilms (Lavrencic *et al.*, 2012).

Biofilms are virulence factors made from complex aggregations of planktonic microorganisms that serve to protect the resident individuals from hostile environments induced by extreme pH, temperature, metals, xenobiotics, antimicrobial agents and even oxygen in some species. Biofilms are covered with a protective layer consisting of a
mixture of extracellular polymeric substances (EPS) secreted by the cells established within the biofilm (Costerton et al., 1999; Joshua et al., 2006).

Biofilm formation requires a concerted mechanism regulated by numerous environmental signals (Karatan and Watnick, 2009) and has been linked to the LuxS enzyme in oral pathogens such as S. mutans, S. intermedius, and Eikenella corrodens (Yoshida et al., 2005; Ahmed et al., 2009; Karim et al., 2013). Recent studies of oral bacteria have provided evidence that LuxS is involved in interspecies signal responses among oral bacteria and consequently may play important roles in the development of virulence and dental biofilms (Shao and Demuth, 2010; Vidal et al., 2011). The LuxS enzyme synthesizes autoinducer 2 (AI-2) as a by-product, which is required for quorum-sensing (QS, cell to cell communication, both inter and intraspecies).

The aims of the experimental procedures performed in this chapter were to:

1. Investigate biofilm production by C. concisus through screening a collection of enteric and oral C. concisus strains for quantitative biofilm formation by crystal violet (CV) assay.
2. Phenotypically characterize the developmental stages of C. concisus biofilms using phase contrast microscopy, confocal laser scanning microscopy (CLSM) and scanning electron microscopy (SEM).
3. Investigate the presence of the potential biofilm related gene (luxS) in the strain collection of oral and intestinal C. concisus strains.
3.2 Materials and Methods

3.2.1 Bacterial strains

*C. concisus* used in this study were 14 intestinal isolates (RCH 3–RCH 7, RCH 11, RCH 12, RCH 15, RCH 19, RCH 20, RCH 23–RCH 26) which were isolated from children suffering from mild to severe bloody diarrhoea using the Cape Town Protocol (Le Roux and Lastovica, 1998) at the Royal Children's Hospital, Melbourne, Australia between June 1993 and June 1995 (Russell, 1995). In addition, 19 oral *C. concisus* (RMIT-O2–RMIT-O15, RMIT-O17, RMIT-O22, RMIT-O23, RMIT-O27 and RMIT-O33) isolated from volunteers from RMIT University, Melbourne, Australia including students, staffs and their families were used in the study. The oral strains were isolated using the Cape Town Protocol (Le Roux and Lastovica, 1998) from healthy individuals by taking their gum swab.

Bacterial strains were grown on HBA in microaerophilic condition and hydrogen enriched environment as explained in Section 2.2.7. The identity of the strains was confirmed by biochemical reactions such as Gram reaction, oxidase production and catalase production following standard methods (Russell, 1995).

Two *C. concisus* type strains *C. concisus* ATCC 51561 and *C. concisus* ATCC 51562 were also used in tests included in this chapter. The reference strains *Campylobacter jejuni* NCTC 11168, *C. jejuni* 81116 (NCTC 11828), *C. coli* NCTC 11366, *C. mucosalis* ATCC 43264, and a non-*Campylobacter* strain *Escherichia coli* ATCC 25922 were all included as control strains in the experiments. *E. coli* was chosen as a negative control and *C. mucosalis* was included because it is closely related to *C. concisus* (On, 1994).
3.2.2 Molecular identification

Molecular identification of *C. concisus* isolates was performed using two sets of primers for *C. concisus* as follows:

### 3.2.2.1 PCR amplification of gyrB of *C. concisus*

All isolates of *C. concisus* used in this study were subjected to PCR amplification of gyrB following the PCR identification protocol described by Matsheka *et al.* (2001) using primer set 1 (Table 3.1). PCR amplification of the *gyrB* in *C. concisus* from boil-prepared chromosomal DNA (Section 2.4.3) was performed at 95°C for 5 min for initial denaturation, 30 cycles of 95°C for 30 sec, 54°C for 30 sec, 72°C for 2 min followed by 72°C for 7 min (final extension) with templates of chromosomal DNA or boiled DNA. The reference strains *C. jejuni* NCTC 11168, *C. jejuni* 81116 (NCTC 11828), *C. coli* NCTC 11366, *C. mucosalis* ATCC 43264, and a non-Campylobacter strain *Escherichia coli* ATCC 25922 were included as negative controls in the PCR. PCR products were visualized by agarose gel (1.5%) electrophoresis.

### 3.2.2.2 PCR amplification of 23S rRNA gene

Genomic DNA was isolated as described in Section 2.4.1 and 2.4.3. All *C. concisus* isolates used in this study were identified by PCR amplification of the 23S rRNA gene using the method described by Istivan *et al.* (2004). In this modified method, two reverse primers (CONI and CON2) were used independently with the common forward primer (MUC1) to group the intestinal and oral isolates. Segment of the 23S rRNA gene was amplified using the primer mix sets 2 and 3 in Table 3.1.
Table 3.1: Primer sequences used in this chapter.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Size</th>
<th>Primer pair</th>
<th>Primer Sequences</th>
<th>Template</th>
<th>Product size (bp)</th>
<th>TA (°C)</th>
<th>Target gene</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pcisus5-F</td>
<td>20</td>
<td>1</td>
<td>AGCAGCATCTATATCAGTT CCCGTTTGATAGGC GATAG</td>
<td>C. concisus gDNA</td>
<td>411</td>
<td>54</td>
<td>gyrB</td>
<td>(Matsheka et al., 2001)</td>
</tr>
<tr>
<td>Pcisus6-R</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MUC1 CON1</td>
<td>19</td>
<td>2</td>
<td>ATGAGTAGCGATAATTGGG CAGTATCGGCAATTCGCT</td>
<td>C. concisus gDNA</td>
<td>306</td>
<td>54</td>
<td>23S rRNA</td>
<td>(Bastyns et al., 1995)</td>
</tr>
<tr>
<td>MUC1 CON2</td>
<td>19</td>
<td>3</td>
<td>ATGAGTAGCGATAATTGGG GACAGTATCAAGGATTTACG</td>
<td>C. concisus gDNA</td>
<td>306</td>
<td>54</td>
<td>23S rRNA</td>
<td>(Bastyns et al., 1995)</td>
</tr>
<tr>
<td>FCCLuxS</td>
<td>22</td>
<td>4</td>
<td>GAAACCATCTAAACGGCAACGG GTCCCATAGCATCAACG TCAAG</td>
<td>C. concisus gDNA</td>
<td>309</td>
<td>55</td>
<td>luxS</td>
<td>This study</td>
</tr>
<tr>
<td>RCCluxS</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
PCR amplification of the 23S rRNA gene in *C. concisus* was performed at 95°C for 5 min for the initial denaturation, 30 cycles of 95°C for 30 sec, 54°C for 30 sec, 72°C for 2 min followed by 72°C for 7 min (final extension). PCR products were visualized by agarose gel (1.5%) electrophoresis. *C. concisus* intestinal and oral isolates were grouped into genomospecies A or B according to their PCR product with either of the reverse primer sets A or B, as compared with the type strains *C. concisus* ATCC 51561 and ATCC 51562.

### 3.2.3 Screening for biofilm-forming *C. concisus* by quantitative crystal violet (CV) assay

All intestinal and oral *C. concisus* isolates were subjected to the screening for biofilm formation by the modified quantitative crystal violet assay (Reeser *et al.*, 2007). *C. concisus* isolates were grown on HBA as explained in Section 2.2.7. A cell suspension with an optical density at 600 nm (A$_{600}$) of 0.5 was prepared in Brucella broth (Bioline, Australia) from HBA plate. A further dilution in Muller Hinton Broth (MHB), Brucella or Columbia (Bioline, Australia) broth was made to reach an OD of ~0.1 containing a cell concentration of $10^8$ CFU/ml which was initially confirmed by a viable plate count. A 96-well tissue culture plate (Corning) was inoculated with 200 µl of the diluted cell suspension (~0.2 X $10^8$ CFU) of *C. concisus*. Plates were incubated at 37°C in hydrogen enriched microaerophilic conditions (6% O$_2$, 10% CO$_2$, 10% H$_2$, and balanced N$_2$) for 24 h, 48 h, 72 h and 96 h.

Following incubation, the medium was removed gently, and the wells were dried for 30 min at 55°C. Then, 0.1 ml of 0.1% crystal violet (CV) solution was added in the wells and the plate was kept for 5 min at room temperature. The unbound CV was
removed by washing twice by immersion in a large beaker containing tap water. The tap water was changed between washes. The plate was dried at 55°C for 15 min. Bound CV in biofilm was solubilized with 80% ethanol-20% acetone and 100 µl of this solution was transferred from the wells and placed into a 96-well tissue culture plate to measure the OD. The absorbance at 570 nm (A_{570}) was determined using a microplate reader (Polarstar Omega microplate reader, BMG Labtech) to indirectly determine the amount of biofilm. The OD reading at 570 nm was interpreted as follows: <0.1: no biofilm, 0.1-0.5: low, 0.5-1.0: moderate, >1.0: high.

Each experiment was performed three times independently and in biological/technical triplicates. In each experiment, a negative control (broth without bacterial inoculum) was included to account for non-specific binding of the CV stain to the 96 well plates. Statistical analysis was done using Mann Whitney test: Two-Sample Assuming Unequal Variances in GraphPad Prism 6.03 for Windows, (GraphPad Software, La Jolla California USA, www.graphpad.com).

3.2.4 Phenotypic characterization of biofilm produced by C. concisus

Following the quantitative assessment of biofilm formation by the CV assay two C. concisus isolates were selected for further investigation of their biofilms by microscopy. Multiple jars were incubated starting at the same time and same conditions for uninterrupted, continuous growth of C. concisus and each time one jar was opened for microscopy. Each experiment was performed three times independently and in biological/technical triplicates.
3.2.4.1 Phase contrast microscopy

Phase contrast microscopy was performed to observe the growth and development of biofilms of *C. concisus* attached to a glass cover slip. Cell suspension of biofilm-producing *C. concisus* was prepared and incubated in hydrogen-enriched microaerophilic condition as described earlier (Section 2.3.1). Plates were incubated for 24, 48, 72, 96 or 120 h to observe the growth and development of biofilms attached to the polystyrene surface phase contrast under 400x magnification using an Olympus CKX41 microscope.

3.2.4.2 Confocal laser scanning microscopy (CLSM)

Confocal laser scanning microscopy was performed to observe arrangement of live and dead cells in *C. concisus* biofilm attached to a glass cover slip. Bacterial suspensions prepared and dispensed into six well flat bottom culture plates as described in Section 2.3.1. After incubation, the biofilms were stained using LIVE/DEAD BacLight bacterial viability kit (Invitrogen) (3.35 μM of Syto-9 and 20 μM of PI) as described in Section 2.3.2. The biofilms were then examined by a Nikon Eclipse Ti inverted microscope equipped with Nikon A1R Fast Laser Scanning Confocal system.

3.2.4.3 Scanning electron microscopy

Scanning electron microscopy was performed to observe the morphology of *C. concisus* growing in biofilm attached to a glass cover slip. Bacterial suspensions were prepared and dispensed into six well flat bottom culture plates as described in Section 2.3.1. After incubation the coverslips were removed very gently from the wells and processed as described in Section 2.3.3. The morphology of biofilms was studied on SEM stubs. A scanning electron microscope (Quanta 200 scanning electron microscope, EM;
FEI Co., Inc., Hillsboro, OR) was used at high vacuum in the secondary electron imaging mode to observe the biofilms at 2500x, 5000x and 20,000x magnifications.

3.2.5 Analysis of luxS in C. concisus genome

A total of 13 C. concisus were selected for analysis of the luxS gene. The sequences of the C. concisus luxS (516 bp) were downloaded from whole genome sequences available in the GenBank, NCBI: C. concisus 13826 (accession number CP000792), C. concisus ATCC 33237T (accession number CP012541), as well as shotgun whole genome sequencing (WGS) C. concisus ATCC 51561 (accession number ANNH00000000), ATCC 51562 (accession number ANNI00000000), UNWCD (AENQ00000000), UNSWCS (ANNG00000000), UNSW1 (ANNF00000000), UNSW2 (ANNJ01000005) and UNSW3 (ANNE00000000). Four luxS sequences were included from the shotgun WGS performed for this PhD project explained in Chapter 5; these were: RMIT-JF1, RMIT-O17, RCH 26 and AUS22-Bd2. A phylogenetic tree was constructed using MEGA7 alignment tool by using the neighbour-joining method for the 13 selected luxS sequences from C. concisus isolates. Multiple alignment was performed using the default parameters which are gap opening penalty 15 and gap extension penalty 6.66.

3.2.6 Molecular detection of luxS from C. concisus

For molecular detection for the presence of luxS in all the isolates of C. concisus a PCR was performed. A new primer set was designed from the sequence of C. concisus 13826 available on the NCBI website. The PCR amplification of the luxS in C. concisus was performed using primer pair 4 (Table 3.1) at 95°C for 5 min for the initial denaturation, then 30 cycles of 95°C for 15 sec, 55°C for 15 sec, 72°C for 10 sec followed by 72°C for 5 min as the final extension with templates of chromosomal DNA or boiled
DNA for 309 bp product. The reference strains *C. jejuni* NCTC 11168, *C. jejuni* 81116 (NCTC 11828), *C. coli* NCTC 11366, *C. mucosalis* ATCC 43264, and a non-*Campylobacter* strain *E. coli* ATCC 25922 was included as negative control in the PCR. PCR products were visualized by an agarose gel (1.5%) electrophoresis.

### 3.3 Results

#### 3.3.1 Molecular identification

##### 3.3.1.1 PCR amplification of gyrB of *C. concisus*

A PCR product of 411 bp was detected in all tested *C. concisus* strains using DNA templates from either clean extracted chromosomal DNA, or from boiled bacterial cells. These results confirm that all *C. concisus* isolates used in this study were successfully identified using both biochemical and molecular methods. No PCR product was detected for other *Campylobacter* spp. strains including the closely related species *C. mucosalis*.

##### 3.3.1.2 PCR amplification of 23S rRNA gene

The identity of *C. concisus* intestinal and oral isolates was confirmed using primers designed to amplify the 23S rRNA gene following the method described by Istivan *et al.* (2004). No PCR product was detected when DNA samples extracted from *C. mucosalis* ATCC 43264, *C. coli* NCTC 11366 or *C. jejuni* 81116 were used as templates. All the intestinal and oral isolates were either positive with primer pair 2 or primer pair 3. On the basis of this difference, they were classified into two genomospecies: A and B. RCH 3–RCH 5, RCH 7, RCH 12, RCH 15, RCH 19 and RCH 20 and the type strain *C. concisus* ATCC 51562 were assigned to genomospecies A; RCH 6 and RCH 11 along with the type strain *C. concisus* ATCC 51561 were assigned to genomospecies B which was published by Istivan *et al.* (2004). RCH 23–RCH 26 produced 306 bp PCR products with primer pair
2 and all of isolates were assigned to genomospecies A. Figure 3.1 is showing some of the isolates producing the expected size of PCR product.

Also 19 oral isolates were identified using 23S rRNA gene following the same method by Istivan et al. (2004). More than half of C. concisus oral isolates (11/19), including RMIT-O2–RMIT-O7, RMIT-O9, RMIT-O10, RMIT-O12, RMIT-O14 and RMIT-O17 produced 306 bp PCR products with primer pair 2 and were assigned to genomospecies A. Eight C. concisus oral isolates (8/19) including RMIT-O8, RMIT-O11, RMIT-O15, RMIT-O22, RMIT-O23, RMIT-O27 and RMIT-O33 produced 306 PCR product with primer pair 3, and were assigned to genomospecies B.

3.3.2 Screening of biofilm-forming C. concisus by Quantitative Crystal Violet (CV) assay

Screening for biofilm-producing C. concisus was performed by the quantitative CV assay with minor modification (Reeser et al., 2007). Biofilm assays were carried out under various conditions to determine the optimum experimental conditions such as optimum inoculum, appropriate media and incubation time for C. concisus. The optimum inoculum for C. concisus was found to be ~10^8 CFU/ml. MHB, Columbia broth and Brucella broth (BB) were used to form biofilm and C. concisus formed better biofilm in Brucella broth. The optimum incubation time was found to be 4 days.
**Figure 3.1:** PCR amplification of genomic DNA from *C. concisus* strains using primer pair 2 (CON1 and MUC1) for 23s rRNA gene sequence of *C. concisus* 13826. The 306 bp PCR products were loaded on a 1.5% (w/v) agarose gel for electrophoresis and then stained with ethidium bromide and visualized on a UV trans-illuminator. Lanes 1-6: PCR product from intestinal *C. concisus* strains: RCH 3, RCH 4, RCH 5, RCH 7, RCH 12, RCH 19; Lane 7: -ve control; Lane 8: +ve control-*C. concisus* ATCC 51562; Lane 9: λ DNA *PstI* digested DNA ladder.
After optimization, both intestinal and oral *C. concisus* were screened for biofilm formation by CV assay. The results of the analysis of intestinal isolates are presented in Figure 3.2. Type strains ATCC 51561 and ATCC 51562 and non-*Concisus* strains *C. jejuni* and *C. coli* were also included to compare the biofilm formation. Among 14 RCH isolates, 11 produced biofilms. RCH 20, 23 and 24 did not form any biofilm. RCH 3, RCH 6, RCH 7, RCH 12, RCH 15, RCH 19, RCH 25 and RCH 26 formed low amount of biofilm. RCH 4 and RCH 11 formed moderate level and only RCH 5 formed high amount of biofilm. Type strains ATCC 51561, ATCC 51562 and *C. jejuni* and *C. coli* produced low amount of biofilms. All of the oral isolates were able to form biofilms. Among 19 oral *C. concisus*, RMIT O2, RMIT-O4, RMIT-O10, RMIT-O11 and RMIT-O27 produced a low amount biofilm. RMIT-O5, RMIT-O6, RMIT-O8 and RMIT-O15 produced moderate levels and RMIT-O3, RMIT-O7, RMIT-O13, RMIT-O17, RMIT-O22, RMIT-O23 and RMIT-O33 produced high levels of biofilms (Figure 3.3). RMIT-O17 was the highest biofilm former. No significant association was found between biofilm formation and the genomospecies typing. Interestingly, the oral isolates were significantly more prolific biofilm producers than the clinical isolates (*p* < 0.05).

### 3.3.3 Phenotypic characterization of *C. concisus* biofilm

Development of *C. concisus* biofilm was observed by phase contrast, CLSM and SEM continuously for 5 days. For this experiment, the highest biofilm producer strain RMIT-O17 was selected. Each experiment was performed three times independently.
Figure 3.2: Crystal violet quantitative assay for screening of biofilm formation by *C. concisus* strains isolated from patients with gastritis. Genomospecies A, Genomospecies B, *C. jejuni* and *C. coli*. The results represent mean values and standard errors of three independent experiments.
Figure 3.3: Crystal violet quantitative assay for screening of biofilm formation by C. concisus strains isolated from oral cavity of healthy persons. Genomospecies A, Genomospecies B. All the oral isolates are abbreviated in the graph with prefix O. The results represent mean values and standard errors of three independent experiments.
3.3.3.1 Phase contrast microscopy

RMIT-O17 was grown in microaerophilic conditions as explained in Section 2.2.7 for 5 days and the biofilm was investigated daily by phase contrast microscopy. Interestingly, several morphological arrangements were observed as the biofilm was growing (Figure 3.4).

After day 1 of incubation, the planktonic bacterial cells are found to be attached at the bottom of the plate (Figure 3.4, A). This is the first stage of biofilm formation, which is attachment. In day 2, the development was observed when cell clusters commenced their development, visualized as multiple cells in contact with one another (Figure 3.4, B). On day 3, it was observed that cell clusters became progressively larger (Figure 3.4, C). On day 4, the multi layered biofilm was observed indicating maturation (Figure 3.4, D). During biofilm maturation in day 4 (or maturation stage), as the majority of the cells are segregated within cell clusters, the cell clusters reach their maximum dimensions, and they become displaced from the edge of the well of the plate. In day 5, cell clusters were observed to undergo alterations in their structure due to the dispersion of bacterial cells from their interior portions (Figure 3.4, E). These bacterial cells were motile and were observed to swim away from the inner portions of the cell cluster through openings in the cluster and enter the bulk liquid. Figure (3.4, E) is an image of a cell cluster taken from the side of a coverslip showing the opening through which the bacterial cells were observed to evacuate the cluster centre. In addition, other cells remaining within the void space were motile. The ability of bacteria to swim freely within the void spaces as observed by microscopy indicated the absence of dense polymers or other gel-like material in the void space.
Figure 3.4: Biofilm development by *C. concisus* RMIT-O17 examined by phase contrast microscopy. Each panel represents each day in biofilm development. (A) Day 1, Bacterial attachment, initial event in biofilm development, bacteria are attached to surface at cell pole (indicated with an arrow). (B) Day 2, cells were cemented to the surface and formed nascent cell clusters (indicated with an arrow). (C) Day 3, cell clusters matured and were several cells thick, embedded in the EPS matrix. (D) Day 4, cell clusters reached maximum thickness. (E) Day 5, cells evacuated interior portions of cell clusters, forming void spaces (indicate with an arrow).
3.3.3.2 **Confocal laser scanning microscopy (CLSM)**

Biofilms of the oral isolate RMIT-O17 were grown for 5 days and different developmental arrangements of biofilm were observed at day 1, 3 and 5 by CLSM under 200x magnifications with a 3.4x zoom. The biofilm was stained with FilmTracer® LIVE/DEAD® Biofilm Viability Kit (Invitrogen) as SYTO 9 was used for live cells and propidium iodide (PI) was used to identify dead cells within the biofilm matrix as shown in Figure 3.5.

After day 1 of incubation, only few green colour emissions were observed from SYTO-9 indicating that only live cells live single cells were attached on the cover slip (Figure 3.5-1). There was no red light emission from PI in day 1 indicated no dead cells on the cover slip (Figure 3.5-1b). This finding of single live cells on the cover slip was in agreement of the finding the results on day 1 from phase contrast microscopy (Figure 3.3A). After 3 days incubation, there were some small aggregates of live cells stained with SYTO-9 (Figure 3.5-2a and 2c). However, still no red colour emission from PI after 3 days incubation indicated only live cells present in the biofilm (Figure 3.5-2b). After 5 days incubation a larger aggregation of cells emitting both red and green colour indicated both live and dead cells in Figure 3.5-3a, b and c. Figure 3.5-3c showing a mature biofilm of *C. concisus* is a mixture of live and dead bacterial cells which supplies the nutrients and protection for the biofilm harbouring bacteria.
Chapter 3: Investigation of *C. concisus* biofilm

**Figure 3.5:** Biofilms of *C. concisus* RMIT-O17 on glass cover slip observed by CLSM at different time of incubation. 1(a, b and c) after 1 day incubation, 2 (a, b and c) after 3 days incubation, 3 (a, b and c) after 5 days incubation; the first column showing stained with SYTO-9 (bright green for live cells) and second column stained with PI (red or orange, or loss of bright green for dead cells). Third column is the combined image of SYTO-9 and PI. In day 1 and 3 no dead cells were found while in day 5 a mixed live and dead cells in the biofilm of *C. concisus.*
3.3.3.3 Scanning electron microscopy (SEM)

Two *C. concisus* strains were selected to observe the biofilm structure by scanning electron microscopy. Oral strain RMIT-O17 was selected as a representative strain of high biofilm-producer and the intestinal strain RCH 26 was selected as a representative moderate biofilm-producer.

The biofilm produced by RMIT-O17 was observed at low magnification (3000x) (Figure 3.6, A) and high magnification (5000x) (Figure 3.6, C). It was found that bacterial cells were attached to the cover slip in aggregates. When RCH 26 was observed at low magnification (2500x) and high magnification (5000x) (Figure 3.6, B & D), bacterial cells were also found to be attached to the cover slip, but in smaller aggregates in comparison to RMIT-O17 aggregates. Interestingly, most of RCH 26 cells were attached to the surface in a planktonic state rather than a biofilm.

Aggregates found within RMIT-O17 biofilm were subjected to SEM with higher magnification (20,000x) and a complex structure of biofilm was observed (Figure 3.7). In the micrograph in this figure, Bacterial cells can be seen to be connected by an extracellular polymeric substance (EPS), which is a defining characteristic of a biofilm. The extracellular material connecting bacteria appears flattened and widespread and bacterial cells were embedded within the EPS (Figure 3.7).

3.3.4 Phylogeny of luxS in *C. concisus*

The alignment *luxS* gene of 13 *C. concisus* strains as states in Section 3.2.7 was performed for phylogenetic analysis. The alignment and the primer pair that was designed to identify *luxS* is shown in Figure 3.8. A phylogenetic tree was generated from this
Figure 3.6: SEM images of *C. concisus* mature biofilms. (A): RMIT-O17 at 3000x magnifications, (B): RCH 26 at 2500x magnifications, (C): RMIT-O17 at 5000x magnifications, (D): RCH 26 at 5000x magnifications showing smaller aggregates (indicated with an arrow) in biofilms of RCH 26 (B, D) compared to RMIT-O17 aggregates (indicated with an arrow) (A, C).
**Figure 3.7:** Biofilm complex produced by RMIT-O17 observed at 20,000x magnification showing the bacterial cells (indicated with arrows) embedded within extracellular material.
alignment using neighbour-joining method is shown in Figure 3.9. The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The tree is drawn to scale, with branch lengths (next to the branches) in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004) and are in the units of the number of base substitutions per site. The analysis involved 13 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 512 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Tamura et al., 2013).

Two clades were generated; one clade contains seven sequences: C. concisus 13826, ATCC 51561, UNSWCD, UNSWCS, UNSW1, UNSW2 and UNSW3. Interestingly all seven sequences in this clade are from genomospecies B (according to PCR of 23S rRNA gene). The other clade contains six sequences: ATCC 33237T, ATCC 51562, RMIT-O17, RMIT-JF1, RCH 26 and AUS22-Bd; these six sequences are all from genomospecies A (according to PCR of 23S rRNA gene).

3.3.5 Molecular detection of luxS from C. concisus

The PCR amplification experiments were carried out with 14 intestinal isolates and 19 oral isolates of C. concisus using the primer set 4 (FCCluxS/RCCluxS in Table 3.1) which was designed to generate a product of the partial luxS from C. concisus (Figure 3.8). Some of the PCR amplification profiles are shown in Figure 3.10. The primer pair successfully amplified products of 309 bp with all the intestinal and oral isolates of C. concisus.
**Figure 3.8:** Nucleotide sequence alignment analysis to design PCR primer pairs for amplification of the *luxS* gene from *C. concisus*. Dots indicate identical bases. Arrows are indicating the forward and the reverse primer used for amplifying *luxS*. Numbers at the left and right refer to base pairs of the full-length *luxS* genes from 13 isolates as stated.
Chapter 3: Investigation of *C. concisus* biofilm

**Figure 3.9:** Phylogenetic analysis of *luxS* genes for the 13 *Campylobacter concisus* strains showing two clades comprised of genomospecies A and genomospecies B. The optimal tree with the sum of branch length = 0.42219599 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths (next to the branches) in the same units as those of the evolutionary distances used to infer the phylogenetic tree.
Figure 3.10: PCR amplification of genomic DNA from *C. concisus* strains using primer pair 4 for the *luxS* gene. The 309 bp PCR products were loaded on a 1.5% (w/v) agarose gel for electrophoresis and then stained with ethidium bromide and visualized on a UV trans-illuminator. Lanes 1-8: intestinal isolates: RCH 3, RCH 4, RCH 5, RCH 6, RCH 7, RCH 11, RCH 12, RCH 16; Lane 9: λ DNA *PstI* DNA ladder; Lanes 10-15: Oral isolate: RMIT-O2, O3, O4, O5, O6, O7.
3.4 Discussion

In spite of the prevalence of \textit{C. concisus} and its pathogenic potential, information on its pathogenic determinants is limited, and molecules possibly involved in biofilm formation have not previously been examined. In this study \textit{C. concisus} isolates from intestinal and oral sources were screened for biofilm formation. PCR of these isolates revealed that all have the \textit{luxS} gene.

All \textit{C. concisus} isolates (intestinal and oral) included in this study were hydrogen-requiring, Gram negative, slightly curved microorganisms. They were positive for oxidase, negative for catalase. The low biochemical activity of \textit{C. concisus} makes it hard for identification by conventional phenotypic techniques (Matsheka \textit{et al.}, 2001). Most laboratory procedures are optimized to isolate the more common species: \textit{C. jejuni} and \textit{C. coli}, which sometimes hinders the proper diagnosis of other \textit{Campylobacter} spp. such as \textit{C. concisus} with additional growth requirements. Molecular methods based on PCR amplification may provide an alternative to culture methods for the detection of \textit{Campylobacter} spp. in clinical specimens. The application of PCR-based assays to detect \textit{Campylobacter} species in clinical and food samples have been reported by different research groups (Maher \textit{et al.}, 2003; Huq \textit{et al.}, 2014). Metagenomic analysis has been successfully used to detect DNA sequences of the \textit{Campylobacter} spp. genome in cases where the standard culture methods were negative (Nakamura \textit{et al.}, 2008).

The identity of all the intestinal and oral isolates of \textit{C. concisus} in this study was confirmed by a well-established highly specific \textit{gyrB} gene PCR described by Matsheka \textit{et al.} (2001). The detection of a 411 bp \textit{C. concisus} species-specific PCR product in all RCH clinical isolates confirmed the strains to be \textit{C. concisus}, while no PCR product was
detected for isolates representing other *Campylobacter* spp. including the closely related species *C. mucosalis* (data not shown). Istivan *et al.* (2004) have used this primer set (Primer set 1 in Table 3.1) to confirm *C. concisus* isolated from stool samples in 2004. Recently, this PCR primer set was used in a multiplex PCR used to detect *Campylobacter* spp., *C. jejuni*, *C. coli* and *C. concisus* (Huq *et al.*, 2014).

In this study all the isolates (intestinal and oral) were classified into two molecular groups using PCR amplification of the 23S rRNA gene following the method of Istivan *et al.* (2004): genomospecies A and genomospecies B. Of the 19 oral isolates, 58% were grouped into the genomospecies A clade, while 42% isolates were grouped into the genomospecies B clade. The majority (87%) of the intestinal isolates, including the type A strain ATCC 51562, were assigned into genomospecies A clade, while only 2 isolates and the type B strain ATCC 51561 (13%) were found in molecular group B (genomospecies B). This finding is in agreement with the results that Istivan *et al.* (2004) had found in their study. However, Aabenhus *et al.* (2005a) and Kalischuk and Inglis (2011) have found genomospecies A predominated in healthy individuals and genomospecies B predominated in diarrheic individuals. Interestingly, these workers identified 3 isolates which were positive for both sets of primers (Table 3.1; primer mix 2 and 3) and one isolate could not be amplified with either primer set. Recently, On *et al.* (2013) attempted to evaluate the distribution of these genomospecies among 47 *C. concisus* isolated from gastroenteritis patients in South African using the same approach, but could not correlate the typing with the clinical outcome.

After confirming the identity of the *C. concisus* isolates by PCR, biofilm production screening was investigated using the quantitative crystal violet assay with
Chapter 3: Investigation of *C. concisus* biofilm

... minor modification (Reeser *et al*., 2007). This assay is widely used assay for detection of biofilm-forming ability of *Campylobacter* spp. and other oral pathogens (Gunther and Chen, 2009; Reuter *et al*., 2010; Domenech *et al*., 2012). To our knowledge, this is the first study comparing biofilm-forming properties of oral and intestinal *C. concisus*. In this study, the crystal violet assay was optimised for *C. concisus* considering various conditions. The optimum inoculum was found as 10^8 CFU/ml for *C. concisus*. Among MHB, Columbia broth and Brucella broth, *C. concisus* produced more biofilm in Brucella broth and matured after 4 days of incubation. Brucella broth (highly nutrient media) is a widely used media for biofilm studies of *Campylobacter jejuni* (Reeser *et al*., 2007; Gunther and Chen, 2009; Joshua *et al*., 2006; Reuter *et al*., 2010). A comparison of the composition of Brucella broth, MHB and Columbia broth it was found that only Brucella broth contains yeast extract, which may promote biofilm production in the liquid state in the laboratory. All the oral isolates were biofilm producers while only three intestinal isolates were found to be non-biofilm producers. Overall, oral isolates produced significantly higher levels of biofilm than the intestinal isolates (*p* <0.05). One possible explanation for forming more biofilm by the oral isolates could be being in an advantageous trait within the oral cavity to escape from toxic oxygen and other adverse conditions. *C. jejuni* NCTC 11168 has been shown to develop biofilm more rapidly under environmental and food-chain-relevant aerobic conditions (20% O_2) than under microaerobic conditions (5% O_2, 10% CO_2) (Reuter *et al*., 2010). Interestingly, the first biofilm study of *C. concisus* by Gunther and Chen (2009) was done with an oral isolate *C. concisus* ATCC 33237^T^ and was found to produce more homogeneous biofilms compared to other *Campylobacter* spp. Recently Lavrencic *et al.* (2012) tested eight *C. concisus* strains for biofilm formation (all from stool samples, including two type strains ATCC 51561 and ATCC 51562) and reported that all eight strains were able to form biofilms.
Chapter 3: Investigation of *C. concisus* biofilm

The biofilm formations by the two type strains (both were isolated from faeces) are in accordance with our study, suggesting that oral *C. concisus* are higher biofilm formers than the intestinal *C. concisus* isolates.

In order to investigate the structure of *C. concisus* biofilms, an oral isolate RMIT-O17 was observed by phase contrast, CLSM and SEM. The biofilm was grown for 5 days and development was observed each day by phase contrast microscopy. Direct observation revealed that the biofilm of *C. concisus* displayed clear developmental stages over the course of the study. This was used to conveniently partition biofilm development into three stages: (i) attachment, (ii) maturation, and (iii) dispersion. The developmental life cycle comes full circle when dispersed biofilm cells revert to the planktonic mode of growth (Figure 3.4). This mode of development in biofilms has been shown previously in *Pseudomonas aeruginosa* (Sauer et al., 2002). In the current study, the mature *C. concisus* biofilm of was found to be a complex matrix of both live and dead cells (Figure 3.5) which is a characteristic of biofilms reported for other organisms. For example, a *P. aeruginosa* artificial biofilm was found to contain 13% dead cells after 8 h of incubation (Takenaka et al., 2001). However, this scenario could vary depending on the type of biofilm and its participating members. *C. jejuni* has been reported to remain culturable for a longer time in a mixed biofilm with *P. aeruginosa* than in monoculture biofilm (Ica et al., 2012). The SEM image of *C. concisus* RMIT-O17 (Figure 3.7) reveals the other important characteristic of the biofilm: the presence of extracellular polymeric substance (EPS). This EPS helps the bacteria to adhere and consists of different chemical components including exo-polysaccharides, proteins, eDNA and other polymers which may provide protection from antibiotics and host defences (Joo and Otto, 2012). The EPS has been identified in biofilms of *C. jejuni* (Joshua et al., 2006), *P. aeruginosa* (Joo and Otto, 2012), and
Borrelia burgdorferi (Sapi et al., 2012). However, no thorough investigation has been conducted to determine the composition of the EPS found in C. concisus biofilms.

LuxS is a signalling protein that has been linked to biofilm formation of many oral pathogens, such as Streptococcus pneumoniae (Vidal et al., 2011), S. mutans (Merritt et al., 2003), and S. intermedius (Ahmed et al., 2009). The luxS gene or its homologue has been identified in C. jejuni, C. coli, C. upsaliensis and C. fetus, however, no luxS gene or homologue was identified in the C. lari genome (Tazumi et al., 2011). Reeser et al. (2007) have found that both quorum-sensing and flagella are required for maximum biofilm formation by C. jejuni. In our study, luxS was successfully amplified from all intestinal and oral C. concisus isolates. To investigate why all the oral isolates were producing biofilm while some intestinal isolates were not producing any biofilm but still harbour the luxS gene, the alignment of the gene was performed (Figure 3.8) and evolutionary tree of luxS was constructed (Figure 3.9). Plummer et al. (2011) analysed luxS gene sequence in C. jejuni 81116 and claimed that a mutation at amino acid position 92 (G to D) in the gene is responsible for the loss of AI-2 activity (approximately 100-fold reduced catalytic activity). In this study 13 C. concisus luxS sequences were analysed and and the amino acid in position 92 was found to be D, which is similar to the wild type C. jejuni 81116. There were some other variations observed in the sequence alignment, and in the resulting phylogenetic tree. Interestingly, these sequences were arranged into two clades which were consistent with the allocation into genomospecies A and B according to 23S rRNA gene. Certainly, there was no correlation between the systemic relationship or the presence of amino acid D at position 92 and amount of biofilm formation. Biofilm formation by C. concisus was significantly associated with the site of isolation.
It can be concluded that *C. concisus* are biofilm-forming bacteria and the ability to form biofilms is linked to the site of isolation. Isolates from the oral cavity were all able to form biofilm *in vitro*, in contrast to only some isolates from the gut. Further, oral isolates produced comparatively higher levels of biofilm than gut isolates. *C. concisus* biofilms were observed to be complex matrices made up of live and dead cells embedded in an EPS. The *luxS* gene was detected in all *C. Concisus* strains, irrespective of biofilm production. Therefore further investigation is required to determine the role of *luxS* in *C. concisus* biofilm formation.
Chapter Four

Knock-Out mutagenesis of the luxS gene in Campylobacter concisus and characterization of ΔluxS mutants

4.1 Introduction

Bacterial cell-to-cell communication is referred to as quorum-sensing, a population-dependent signalling mechanism that involves the production and detection of extracellular signalling molecules (Williams, 2007). Quorum-sensing in many Gram negative bacteria is based upon the signalling molecule homoserine lactone (HSL) which is classified as an autoinducer-1 (AI-1). Furthermore, an alternative quorum-sensing system in Gram negative and positive bacteria is mediated by furanosyl borate diester referred as AI-2. LuxS protein is responsible for synthesis of AI-2 precursor, DPD (Williams, 2007). LuxS has received intense interest because it plays an important role in the development and spatial organization of oral biofilms (Pecharki et al., 2008; Vidal et al., 2011; Wang et al., 2013). LuxS appears to be essential for the development of Porphyromonas gingivalis–Streptococcus gordonii mixed-species biofilms (McNab et al., 2003). There is also evidence that inactivation of luxS effects biofilm formation of Streptococcus mutans (Merritt et al., 2003; Yoshida et al., 2005). The presence of luxS in Campylobacter spp. has been reported (Reeser et al., 2007; Plummer et al., 2011). Reeser et al. (2007) reported that C. jejuni flaAB and luxS mutants produced a significantly reduced amount of biofilm in comparison to that of wild type strains. The luxS gene or its homologue was detected in C. jejuni, C. coli, C. upsaliensis and C. fetus by PCR or by southern blotting hybridization, but not in the C. lari genome (Tazumi et al., 2011). To our knowledge, no study has been performed yet in C. concisus to investigate luxS and its
function on biofilm formation and other virulence factors such as invasion, adhesion or motility.

In Chapter 3 the luxS gene was detected in all oral and intestinal C. concisus isolates, and the phenotypic characteristics of C. concisus biofilms was investigated. The objective of this chapter was to investigate the role of luxS in biofilm formation and in bacterial interaction with host cells. This aim was achieved by:

2. Creating luxS knock-out strains of C. concisus.
3. Testing the ability of the mutant to form a biofilm.
4. Testing the invasiveness and adhesive properties of the mutant strain on in vitro in an intestinal cell line INT 407.
5. Creating a complementary luxS strain to prove that the observed effects were a result of luxS expression.

4.2 Materials and Methods

4.2.1 Primer design

The primer pair 4 described in Chapter 3 Table 3.1 was employed for the detection of luxS in the chromosomal DNA. A new set of degenerate primers were designed from the flanking regions of luxS to amplify the whole luxS gene (516 bp) including the promoter region. To design the primers, all the available whole genome sequences (WGS) of C. concisus (C. concisus 13826 and ATCC 33237; shotgun WGS ATCC 51561, ATCC 51562, UNWCD, UNSWCS, UNSW1, UNSW2 and UNSW3) were analysed and a degenerate forward primer and two reverse primers were designed.
The primers used in this study are included in the Table 4.1. All primers were designed using Clone Manager software (version 7.11).

4.2.2 Bacteria and plasmids

The bacterial strain *C. concisus* RMIT-O17 was selected to create a *luxS* null mutant by inserting a kanamycin marker. The bacteria and plasmids used in this study to create a *luxS* mutant are listed below:

- **Bacterial strains**
  - *C. concisus* RMIT-O17
  - *E. coli* DH5α

- **Plasmids**

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBluescript SKII</td>
<td>cloning vector pBluescript SKII</td>
</tr>
<tr>
<td>pMW2</td>
<td>vector with ampicillin and kanamycin resistance cassettes</td>
</tr>
<tr>
<td>pBK <em>luxS</em></td>
<td>A 1,380 bp PCR product of the <em>luxS</em> from <em>C. concisus</em> RMIT-O17 in pBluescript SKII</td>
</tr>
<tr>
<td>pBK <em>luxS</em> kan</td>
<td>pBK <em>luxS</em> containing a kanamycin cassette</td>
</tr>
<tr>
<td>p3316</td>
<td>Plasmid present in RCH 26</td>
</tr>
<tr>
<td>pBC SK +</td>
<td>cloning vector with chloramphenicol resistant cassette</td>
</tr>
<tr>
<td>pBC <em>luxS</em></td>
<td>A 1,380 bp PCR product of the <em>luxS</em> from <em>C. concisus</em> RMIT-O17 in pBC SK +</td>
</tr>
</tbody>
</table>
p3316 BC *luxS* shuttle vector for *E. coli* and *C. concisus* containing *luxS* and p3316 from RCH 26

4.2.3 Genomic DNA extraction

Genomic DNA extraction from the bacterial cells was performed by the method described in Chapter 2 Section 2.3.1 using the Wizard Genomic DNA Purification Kit (Promega, Australia). The extracted DNA was dissolved in 50 μl of dilution buffer in a micro-centrifuge tube (1.5 ml). The concentration and purity of gDNA was estimated using a Biophotometer (Eppendorf, Australia) by measuring the optical density (OD) at 260 nm and the ratio of OD obtained at 260 nm and 280 nm respectively. The extracted DNA was stored at –20°C.

4.2.4 PCR optimization

The *luxS* primers (primer pair 1 and 2a, 2b in Table 4.1) designed using the Clone Manager Suite 7.11 (Sci Ed Central) were used to amplify *luxS* from *C. concisus* RMIT-O17 gDNA. Clone Manager was used to determine the optimal annealing temperature for the designed primers. MyTaq red mix kit (Bioline, Australia) was used to amplify the gene. The PCR products were visualised in 1.5% agarose gel as described in Chapter 2 Section 2.4.5.3.
Table 4.1: The primers used in this study are listed (5’-3’).

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Size (bp)</th>
<th>Primer pair</th>
<th>Primer Sequences</th>
<th>Template</th>
<th>Product size (bp)</th>
<th>$T_A$ (°C)</th>
<th>Target gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>luxScfc</td>
<td>30</td>
<td>1</td>
<td>GATCGGTACCATGAGCCTTCTTGCRGTRTC</td>
<td>RMIT-O17</td>
<td>1380</td>
<td>52</td>
<td>luxS</td>
</tr>
<tr>
<td>luxScr1</td>
<td>30</td>
<td>2a</td>
<td>GCTAGAGCTCATAGAAGCGGCTCGTGAGG</td>
<td>luxScr2</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>luxScr2</td>
<td>31</td>
<td>2b</td>
<td>GCTAGAGCTCCAAGTCTCGACGCCTAGAAG</td>
<td>luxScr1</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inv F O17</td>
<td>28</td>
<td>3</td>
<td>CGTAGGATCCGCCCATCGGTAGATGTC</td>
<td>pBK luxS</td>
<td>3400</td>
<td>54</td>
<td>pBluescript containing luxS</td>
</tr>
<tr>
<td>Inv R O17</td>
<td>33</td>
<td></td>
<td>GAGCGGTACCTCTACGCGGTGCGGTGAGATG</td>
<td>luxS</td>
<td>3400</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>luxSF com</td>
<td>30</td>
<td>4</td>
<td>GGTCTCTAGAATGAGCCTTTCTTGCRRGTRTC</td>
<td>RMIT-O17</td>
<td>1380</td>
<td>52</td>
<td>luxS</td>
</tr>
<tr>
<td>luxSR com</td>
<td>31</td>
<td></td>
<td>GCTAGAGCTCCAAGTCTCGACGCCTAGAAG</td>
<td>luxS</td>
<td>3400</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>p3316 inv F</td>
<td>28</td>
<td>5</td>
<td>GCTAGGATCCCGAAAGGCCGHTGATTG</td>
<td>RCH 26</td>
<td>3300</td>
<td>55</td>
<td>RCH 26 Plasmid</td>
</tr>
<tr>
<td>p3316 inv R</td>
<td>32</td>
<td></td>
<td>GTGACTGCAGCCCTTTGGCTTTAACCCTCTACC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Primers were named according to the gene’s name with a suffix letter ‘f’ or “r”. The letter “f” indicates forward primer and the letter “r” indicates reverse primer and R = Ag
Chapter 4: Knock-out of luxS in C. concisus

4.2.5 Cloning of PCR product into primary vector (pCR 2.1 TOPO)

The amplified product was cloned into pCR 2.1-TOPO vector using TOPO TA Cloning Kit (Invitrogen). For the cloning reaction, 1 μl fresh amplified product was added to 1μl salt solution (final concentration of 200 mM NaCl and 10 mM MgCl₂), 3 μl molecular grade water, and 1 μl pCR 2.1-TOPO vector (10 ng) in a sterile tube. The tube was mixed gently and incubated for 5-10 min at room temperature, and placed on ice. The “ligated” product was transformed into electrocompetent E. coli DH5α cells. A range of 1-3 μl of the ligated product was mixed with 50 μl of electrocompetent E. coli DH5α cells and the mixture was transferred to the bottom of an ice-cold cuvette (with a 2 mm gap). The Gene Pulser (Bio-Rad) instrument was used to perform electroporation at 200 Ω, 25 μF capacitance, and 2.0-2.5 kV. One ml of SOC medium was added to the cuvette to resuspend the cells and incubated for 1 h at 37°C. The “transformants” were plated out onto a pre-warmed LB agar plate containing 100 μg/ml ampicillin and X-gal (40 μl of 40 mg/ml solution). The “concentrated transformants” were also plated onto a second LB agar plate containing ampicillin and X-gal (40 μl of 40 mg/ml solution), by resuspending the pellets in a small volume of SOC medium after centrifugation for 3 min. The plates were incubated overnight at 37°C (protected from light).

After overnight incubation, the plates were analysed for blue/white colonies. Several white colonies were selected. Each isolated colony was inoculated into 5 ml LB broth containing 100 μg/ml ampicillin followed by overnight incubation at 37°C with shaking at 200 rpm. Plasmid DNA was extracted from overnight culture using the ISOLATE II Plasmid kit (Bioline) following the manufacturer’s instructions. The isolated plasmid DNA was stored at -20°C. Plasmid DNA integrity was confirmed by restriction enzyme digestion and sequencing reaction was performed on the samples.
The restriction enzyme EcoRI (New England Biolabs) was used to digest the plasmid DNA samples. For the digestion reaction, 2 μl plasmid DNA was mixed with 1 μl 10x Digestion buffer H (New England Biolabs), 0.5 μl EcoRI (5 U), and made up to a final volume of 10 μl by adding 6.5 μl molecular grade water in micro-centrifuge tube. The reagents were mixed gently in the micro-centrifuge tube and given a short spin and incubated at 37°C for 1 h. The entire digested samples were loaded onto 1.5% agarose gels, and separated by agarose gel electrophoresis.

The insert was confirmed by sequencing of the plasmid pCR 2.1-TOPO with the insert using M13 forward and M13 reverse primers as described previously in Section 2.3.5.5.

4.2.4 Cloning of luxS into pBluescript vector

4.2.4.1 Isolation of pBluescript vector

The pBluescript vector was collected from School of Sciences, RMIT University central stocks and was used for the experiments. The empty pBluescript vector in E. coli DH5α was inoculated into 5 ml LB broth containing ampicillin (100 μg/ml), and incubated overnight at 37°C with vigorous shaking (225 rpm). The plasmid was extracted from the culture using ISOLATE II Plasmid kit (Bioline) following the manufacturer’s instructions. The plasmid map of vector pBluescript is shown in Appendix II.

4.2.4.2 Double digestion of pBluescript vector with KpnI and SacI

The “empty” pBluescript vector was digested with KpnI and SacI. Each double digestion reaction was performed by adding 5 μl of plasmid DNA to 0.5 μl KpnI (5U), 0.5 μl SacI (5U), 2 μl 10x Multi-Core buffer and made up to a final volume of 20 μl by adding
11 µl molecular grade water in a PCR tube. The reaction tubes were gently mixed and incubated at 37°C water bath for 1 h. The entire digests were loaded onto 1.5% gel, and separated by gel electrophoresis for 2 h at 40 V.

4.2.4.3 **Excision and gel extraction of double-digested DNA bands**

The DNA bands were excised using a scalpel blade and purified as explained in Section 2.3.5.4 using the ISOLATE II Gel and PCR kit (Bioline) as per the manufacturer’s instructions. The samples were eluted in Elution Buffer, and stored at -20°C.

4.2.4.4 **Ligation of inserts into linearised pBluescript vector**

For the ligation reaction, ~50 ng of vector DNA was mixed with one- to fivefold molar excess of the target DNA fragment (Sambrook and Russell, 2001). The required amount was calculated from the formula:

\[
\text{ng of DNA insert} = \frac{\text{ng vector} \times \text{size of insert (kb)}}{\text{vector size of vector (kb)}} \times \text{molar ratio of insert to vector}
\]

The ligation reaction was prepared in a PCR tube and incubated overnight at 14°C in a water bath. The ligated products were stored at -20°C. Ligated products were confirmed by loading the samples on to gel and separated electrophoretically for 60 min at 100 V.

4.2.4.5 **Electro-transformation of ligated products**

Ligated product was electro-transformed into electrocompetent *E. coli* DH5α as described in Section 2.4.7.3.
4.2.4.6 Confirmation of successful electro-transformation by PCR and restriction enzyme digestion

The colonies were inoculated into LB broth with appropriate antibiotics and incubated overnight at 37°C by shaking (200 rpm). The plasmid DNA was extracted using the ISOLATE II Plasmid kit (Bioline) and analysed by restriction enzyme analysis. PCR was performed to confirm the product using the specific primers.

4.2.7 Creation of C. concisus ΔluxS mutant strain

4.2.4.7 Inverse PCR

Inverse PCR (IPCR) is a method used to amplify and clone unknown DNA that flanks one end of a known DNA sequence (Sambrook and Russell, 2001). The unknown sequence is amplified by two primers that bind specifically to the known sequence and point in opposite directions. Therefore the entire plasmid is amplified. IPCR was performed using primer pair 3 (Table 4.1). The plasmid pMW2, collected from School of Sciences, RMIT University stocks was grown at 37°C overnight in E. coli DH5α in LB broth with kanamycin antibiotic (50 µg/ml). The kanamycin cassette was extracted from the purified plasmid by restriction enzyme digestion with BamHI. The plasmid map of vector pMW2 is given in Appendix III.

IPCR was used to delete part of the gene of interest (136 bp to change the reading frame) and to introduce a BamHI site. The purified PCR product (purified using ISOLATE II Gel and PCR kit, Bioline) and the empty pMW2 vector were digested with BamHI. The entire digests were loaded onto 1.5% gel, and separated by gel electrophoresis for 2 h at 40 V. The DNA bands were excised using a scalpel blade and purified using the ISOLATE II Gel and PCR kit (Bioline) as per the manufacturer’s instructions. The samples were eluted...
in Elution Buffer, and stored at -20°C. Then the products were ligated with the 1.4 Kb Kanamycin cassette of the pMW2 vector. The ligated products were electro-transformed into *E. coli* DH5α cells as described previously in Section 2.4.7.3.

### 4.2.4.8 Confirmation of ligation by Restriction enzyme digestion

Colonies were picked from plates and the plasmid DNA was isolated. The restriction enzymes *Kpn*I and *Sac*I were used to digest plasmid DNA samples. The entire digested samples were loaded onto 1.5% agarose gels, and separated by agarose gel electrophoresis.

### 4.2.4.9 Confirmation by DNA sequencing

The accuracy of the nucleotide sequence of the insert was confirmed by sequencing of the plasmid pBluescript with the insert using M13 forward and M13 reverse primers as described previously in Section 2.4.5.5.

### 4.3 Transformation into *C. concisus*

Transformation of the pBK *luxS*, pBluescript plasmid containing ‘*luxS* with the kanamycin cassette’ was performed by either natural transformation or electroporation.

#### 4.3.1 Natural transformation of *C. concisus*

*C. concisus* RMIT-O17 was grown on HBA in an incubator at 37°C with hydrogen gas mix. Transformation was performed either on an agar surface or in a biphasic system. For transformation on MHA, recipient cells of *C. concisus* were grown on MHA for 24 h in microaerophilic condition. Cells were collected after 24 h in MHB and then spread on a fresh MHA plate at about 5 x 10⁸ cells per plate and incubated for 6 h in microaerophilic conditions.
condition. After incubation aliquots of DNA (10 µg in TE buffer) were spotted directly onto the inoculated agar surface without additional mixing or spreading, and incubation was continued for 5 h.

For transformation in a biphasic system, fresh recipient cell suspensions of (1 x 10^8 to 5 x 10^7 cells per ml of MHB) were transferred (0.2 ml per tube) to test tubes (10 by 120 mm) containing 1.5 ml MHA and incubated for 2 to 6 h. DNA samples (10 µg in TE buffer) were added, and incubation was continued for 3 to 5 h.

The cells were harvested in 1 ml of MHB from MHA plate or test tubes plated out on HBA plates containing kanamycin. The plate was incubated in microaerophilic conditions for 3 d at 37°C. The colonies were confirmed by colony PCR.

4.3.2 Electroporation of C. concisus

C. concisus RMIT-O17 cells were grown on HBA plates for 16 h at 37°C at microaerophilic conditions. Bacterial colonies were harvested using 2 ml Muller Hinton broth or wash buffer. Cells were pelleted by centrifugation (>1000 x g for 5 min) and the washing steps were repeated. Finally, cells were resuspended in ice cold wash buffer and immediately stored in 50 µl or 100 µl aliquots at -80°C (Section 2.4.8.1). Frozen competent cells and the cuvettes were kept on ice. A range of 5-10 µg of DNA was mixed with 50 µl or 100 µl aliquots of competent cells and electroporation was performed at 2.5 kV, 200 Ω, 25 µF in a 2 mm gap cuvette. A total of 100 µl of SOC medium was added and gently spread onto non-selective media. The plates were incubated 37°C for 5 h under microaerophilic conditions. The cells were harvested in 1 ml of MH Broth and plated out.
on HBA plates containing kanamycin. Plates were incubated at 37°C for 3 d. The colonies were confirmed by colony PCR.

4.4 Genetic complementation of ΔluxS mutant strain

4.4.1 Construction of a shuttle vector

A chloramphenicol resistant plasmid, pBC SK + was selected for this experiment (Appendix III). Primer pair 4 (Table 4.1) were used to amplify the luxS gene from RMIT-O17 while a different restriction site (XbaI) was inserted instead of KpnI in the forward primer for ease of manipulation of the vector pBC SK +. The luxS gene was inserted into pBC SK + by ligation and cloned into E. coli DH 5α as described in Section 4.3. The new constructed plasmid was named as pBC luxS.

To construct the E. coli-C. concisus shuttle vector a small plasmid of 3316 bp size isolated from RCH 26 carrying mob, repC, and repE and condensin complex subunit 2 was amplified with inverse PCR primers inserting BamHI and PstI restriction enzymes sites. An attempt was taken to ligate the IPCR product with pBC luxS. As a result a new shuttle vector would be constructed named as p3316 BC luxS (Figure 4.1) and it could be maintained in E. coli DH5α.
Figure 4.1: Predicted genetic map of the shuttle vector p3316 BC *luxS* for complementation.
4.4.2 Transformation of shuttle vector p3316 BC luxS

Constructed shuttle vector p3316 BC luxS DNA prepared from E. coli DH5α transformants will be used to transform to mutant RMIT-O17 to complement the knocked out gene by natural or electroporation as described in Section 4.3.

4.5 Phenotypic characterization of ΔluxS-O17

The wild type strain RMIT-O17 and the mutant strain ΔluxS-O17 were compared by the following characteristics:

4.5.1 Quantitative Crystal violet Assay

The wild type strain RMIT-O17 and the mutant strain ΔluxS-O17 were tested for biofilm formation by modified quantitative crystal violet (CV) assay (Reeser et al., 2007) following the procedure explained in Section 3.2.3. Each experiment was performed three times independently and in biological/technical triplicates. A negative control (broth without bacterial inoculum) was included to account for non-specific binding of the CV stain to the 96 well plates. Statistical analysis was done by t-test: Two-Sample Assuming Unequal Variances by using GraphPad Prism 6 statistical tools. For all statistical analyses, a confidence level of 95% was defined.

4.5.2 Motility assay

The motility of C. concisus strains were investigated at 37°C on semi-solid agar plates using a modification of a method described by Adler et al. (2014). Either semi-solid BB containing 0.4% agar (BBA 0.4%) or MH containing 0.4% agar (MHA 0.4%) were used for this test. Overnight cultures of C. concisus strains were adjusted to 10⁸ CFU/ml and 3.4 µl of this suspension were dropped on semi-solid BBA or MHA. After 72 h
incubation at 37°C in microaerophilic hydrogen enriched condition, the diameters of the growth were measured. Growths of mutants were normalized to the wild type diameter (100%). Results reported are the median of six independent assays.

4.5.3 Adhesion and Invasion assay

Adhesion and invasion of the wild type strain RMIT-O17 and the mutant strain ΔluxS-O17 were carried out in intestinal cell line INT 407 cells based on the methods previously described in Section 2.5.2.4. Each assay were performed in triplicates and repeated three times on different days. Statistical analysis was done by t-Test: Two-Sample Assuming Unequal Variances by using GraphPad Prism 6 statistical tools. Invasive index was calculated using the formula described by (Larson et al., 2008):

Results are expressed as a percentage of the total number of colonies adhered or invaded divided by the inoculum.

\[
\text{Total % Adhered } C. \text{ concisus} = \frac{\text{Number of colonies adhered}}{\text{Original inoculum}} \times \text{dilution factor} \times 100
\]

\[
\text{Total % invaded } C. \text{ concisus} = \frac{\text{Number of colonies invaded}}{\text{Original inoculum}} \times \text{dilution factor} \times 100
\]

Further, calculate % of \( C. \text{ concisus} \) cells invaded compared to those that adhered using the total number of cells adhered.

\[
\text{Invasive index} = \frac{\text{Number of invaded } C. \text{ concisus}}{\text{Number of adherent } C. \text{ concisus}} \times 100
\]
4.5.4 Phenotypic characterization of biofilm by microscopy

Biofilm produced by the wild type strain RMIT-O17 and the mutant strain $\Delta$luxS-O17 were observed and compared using confocal laser scanning microscopy (CLSM) and scanning electron microscopy (SEM) as described in previous Section 2.3.2 and 2.3.3.

4.6 Results

4.6.1 Creation of C. concisus $\Delta$luxS mutant

*C. concisus* RMIT-O17 was selected for generation of the luxS gene knock-out. To design the new set of primers a total of nine WGS of *C. concisus* available in the database were analysed as the WGS of RMIT-O17 was not available at that time. Interestingly, a difference was found in the flanking genes of luxS in *C. concisus* genomes. The upstream gene of the *luxS* was the same (Adenine-specific DNA methyltransferase) in all *C. concisus* genomes but the gene located downstream of *luxS* in the WGS was different as shown in Table 4.2. *C. concisus* 13826, ATCC 51561, UNSWCD, UNSW1 UNSW2 and UNSW3 had a signal transduction protein following *luxS* in the genome and all these strains are from genomospecies B. On the other hand, ATCC 51562 and ATCC 33237T had a putative protein containing molybdenum cofactor (MOCO) sulfurase C-terminal (MOSC) domain are from genomospecies A. However, the UNSWCS genome had the same molybdenum cofactor (MOCO) sulfurase C-terminal (MOSC) domain following *luxS* gene although this strain belongs to genomospecies B. The *luxS* gene, including the flanking regions from RMIT-O17 was successfully amplified and cloned into the TOPO pCR 2.1 plasmid using primer pair 1 and 2b (Table 4.1). The inserted product was sequenced and aligned with the available reference *C. concisus* 13826 *luxS*. LuxS from RMIT-O17 had 96% similarity with LuxS from *C. concisus* 13826 (Figure 4.2).
<table>
<thead>
<tr>
<th></th>
<th>Sequence 1</th>
<th>Sequence 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>LuxS_017</td>
<td>MPLLDSFCVDHVKMQAPGVR1LSMKTPKGDDISVFDLRCNPNEEILPEKGTHTLEHLF</td>
<td>60</td>
</tr>
<tr>
<td>LuxS_13826</td>
<td>MPLLDSFCVDHVKMKAPGVR1LSMKTPKGDDISVFDLRFCPNEEILPEKGTHTLEHLF</td>
<td>60</td>
</tr>
<tr>
<td>LuxS_017</td>
<td>AGFRMRNHNLGVEIDISPSMCRTGRTFYMTPNEEAVKKAWLASMKDIEVXQDKI</td>
<td>120</td>
</tr>
<tr>
<td>LuxS_13826</td>
<td>AGFRMRNHNLGVEIDISPSMCRTGRTFYMTPSEEAVKKAWLASMKDILEVXQDKI</td>
<td>120</td>
</tr>
<tr>
<td>LuxS_017</td>
<td>PELNKFCQGTYKMKSLDEHAHAKILAQGLVIIINDEIKLDVMDLKKH171</td>
<td></td>
</tr>
<tr>
<td>LuxS_13826</td>
<td>PELNKFCQGTYKMKSLDEHAHAIASKILAQGLVIIINDEIKLDVMDLKKH171</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4.2:** Alignment of RMIT-O17 LuxS with *C. concisus* 13826 LuxS protein using the CLUSTAL O algorithm (1.2.2) (www.ebi.ac.uk). The amino acid residues that are conserved in two sequences of the alignment are marked by *.
**Chapter 4: Knock-out of luxS in C. concisus**

**Table 4.2:** luxS and the flanking genes of nine C. concisus strains available in NCBI database and their genomospecies.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Flanking Genes of luxS</th>
<th>Genomospecies</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. concisus 13826</td>
<td>Adenine-specific DNA methyltransferase Upstream, Cache1 sensor-containing MCP-domain signal transduction protein Downstream</td>
<td>B</td>
<td>CP000792</td>
</tr>
<tr>
<td>ATCC 33237T</td>
<td>Adenine-specific DNA methyltransferase</td>
<td>A</td>
<td>CP012541</td>
</tr>
<tr>
<td>ATCC 51561</td>
<td>Adenine-specific DNA methyltransferase Cache1 sensor-containing MCP-domain signal transduction protein</td>
<td>B</td>
<td>ANNH00000000</td>
</tr>
<tr>
<td>ATCC 51562</td>
<td>Adenine-specific DNA methyltransferase</td>
<td>A</td>
<td>ANNI00000000</td>
</tr>
<tr>
<td>UNSWCD</td>
<td>Adenine-specific DNA methyltransferase</td>
<td>B</td>
<td>AENQ00000000</td>
</tr>
<tr>
<td>UNSWCS</td>
<td>Adenine-specific DNA methyltransferase</td>
<td>B</td>
<td>ANNG00000000</td>
</tr>
<tr>
<td>UNSW1</td>
<td>Adenine-specific DNA methyltransferase</td>
<td>B</td>
<td>ANNF00000000</td>
</tr>
<tr>
<td>UNSW2</td>
<td>Adenine-specific DNA methyltransferase</td>
<td>B</td>
<td>ANNJ00000000</td>
</tr>
<tr>
<td>UNSW3</td>
<td>Adenine-specific DNA methyltransferase</td>
<td>B</td>
<td>ANNE00000000</td>
</tr>
</tbody>
</table>
The scheme for creating the knock-out ΔluxS mutant is shown in Figure 4.3 and the full construction for ΔluxS mutant is shown in Figures 4.4-4.9.
Figure 4.3: Flow chart showing the strategy adopted in creating knock-out of luxS in RMIT-O17. Plasmid pCR 2.1 containing cloned fragments of luxS gene including flaking region of RMIT-O17 were cut using SacI and KpnI restriction enzymes and inserted in pBluescript II SK (+). The resulting plasmid pBK luxS was linearized and BamHI restriction sites were inserted by inverse PCR. The kanamycin cassette was cut from pMW2 using BamHI and then ligated into a linearized pBK luxS backbone. The resulting construct pBK luxS kan was transformed into RMIT-O17 for double crossover.
Chapter 4: Knock-out of \textit{luxS} in \textit{C. concisus}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure4.4.png}
\caption{Amplification of flanking regions of \textit{luxS}, a 1,380 bp PCR product from \textit{C. concisus} RMIT-O17. Lane 1 & 2: PCR product from flanking regions of \textit{luxS}; Lane 3: Negative control; Lane 4: \textit{\lambda} DNA \textit{PstI} DNA ladder.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure4.5.png}
\caption{Digestion of pCR 2.1-TOPO/ \textit{luxS} with \textit{EcoRI} reveals fragments at the expected sizes of 3.9 kb and 1,380 bp. Lane 1: pCR 2.1-TOPO (3.9 kbp) and 1,380 bp PCR product; Lane 2: \textit{\lambda} DNA \textit{PstI} DNA ladder.}
\end{figure}
Figure 4.6: Digestion of pBK luxS by restriction enzyme KpnI and SacI produced fragments at the expected sizes of 2.9 kb and 1,380 bp. Lane 1: pBK luxS digested products by KpnI and SacI; Lane 2: λ DNA PstI DNA ladder.

Figure 4.7: Inverse PCR of pBK luxS produced a 4.3 kb PCR product. Lane 1: Inverse PCR product of pBK luxS; Lane 2: Negative control; Lane 3: λ DNA PstI DNA ladder.
**Figure 4.8:** Confirmation of kanamycin insert in pBK luxS kan by restriction enzyme digestion with *BamHI*/*KpnI* and *SacI*. Lane 1: digested products of 4 kb and 1.3 kb from pBK luxS kan by *BamHI* digestion; Lane 2: digested products of 2.8 kb and 2.7 kb from pBK luxS kan by *KpnI* and *SacI* digestion; Lane 3: λ DNA *PstI* DNA ladder.

**Figure 4.9:** PCR showing the predicted size of luxS knockout in pBK luxS kan and the size of wild type luxS. Lane 1: λ DNA *PstI* DNA ladder; Lane 2-3: Amplicon at predicted size (1.7 kb) of knock-out luxS (knock-out); Lane 4: luxS wild type (309 bp).
4.6.2 Genetic complementation

To complement ΔluxS-O17, an attempt to insert an active second copy of luxS by an E. coli-C. concisus shuttle vector within the mutant strain was performed. The genetic complementation experiment could not be completed successfully due to limited time. The luxS from RMIT-O17 was amplified with the new primer set and successfully inserted in the multiple cloning site (MCS) by XbaI and SacI restriction site in plasmid pBC SK+. The inverse PCR from a plasmid in RCH 26 was performed by inserting the same restriction sites but unfortunately, the second ligation of this new plasmid and the plasmid isolated from RCH 26 was not successful. The results of the initial steps are described in Figures 4.10-4.12.

4.6.3 Phenotypic characterization of the ΔluxS mutant

4.6.3.1 Quantitative crystal violet (CV) assay

Biofilm formation by the wild type RMIT-O17 and ΔluxS-O17 was quantitated by CV assay as previously described in Section 3.2.3. The result is presented in Figure 4.13. The wild type produced a significantly higher amount of biofilm compared to the ΔluxS-O17 (p < 0.05) (Unpaired t-test). In each experiment a negative control (no bacteria) was included to account for non-specific binding of the stain.

4.6.3.2 Motility assay

Motility of ΔluxS-O17 was normalized to the wild type RMIT-O17 (100%). The results are presented in Figure 4.14. ΔluxS-O17 showed reduced motility (approximately 70% of wild type RMIT-O17 in BBA) (p < 0.02). Mean diameter of RMIT-O17 was 8.8 mm and ΔluxS-O17 was 6.1 mm in semi-solid BBA.
Chapter 4: Knock-out of luxS in *C. concisus*

**Figure 4.10:** Optimized PCR amplification of flanking regions of luxS, a 1380 bp PCR product from RMIT-O17. Lane 1: PCR product from flanking regions of luxS; Lane 2: Negative control; Lane 3: λ DNA *Pst*I DNA ladder.

**Figure 4.11:** Digestion of pBC luxS by restriction enzymes *Xba*I and *Sac*I produced fragment at the expected sizes of 3.4 kb and 1,380 bp. Lane 1: λ DNA *Pst*I DNA ladder; Lane 2: pBC luxS digested by *Xba*I and *Sac*I.
**Figure 4.12:** Inverse PCR for p3316 in RCH 26 produced a 3.3 kb PCR product. Lane 1: Inverse PCR performed on RCH 26 gDNA for p3316; Lane 2: negative control; Lane 3: λ DNA PstI DNA ladder.
Figure 4.13: Reduced biofilm formation by *C. concisus* luxS mutant. RMIT-O17 and ΔluxS-O17 were subjected to a standard biofilm CV assay. Biofilm formation was measured after 4 days. Red bar represents RMIT-O17, striped red bar represents the ΔluxS-O17 mutant. ΔluxS-O17 showed significantly reduced motility (*p* < 0.02) (by t-test). Experiments were performed three times in triplicate, and error bars represent standard standard error of the mean.
Figure 4.14: Reduced motility by *C. concisus* luxS mutant. RMIT-O17 and ΔluxS-O17 were subjected to motility assay on semi-solid agar (0.4% BBA). The growth diameter was measured after 72 h incubation on semi solid agar. Red bar represents RMIT-O17, striped red bar represents the ΔluxS-O17 mutant. Experiments were performed six times independently, and error bars represent standard error of the mean.


4.6.3.3 Adhesion and invasion assay

The adhesion and invasion assay was performed for wild type RMIT-O17 and mutant ΔluxS-O17 on the INT 407 epithelial cell line. The experiment was performed at a multiplicity of infection (MOI) of 100 as described previously in Section 2.5.2.4. The proportion of *C. concisus* adherent or invading in the cell line was calculated (Figure 4.15 and 4.16). Invasive index was calculated using the formula described by Larson *et al.* (2008) and presented in Figure 4.17.

There was no statistically significant difference in the proportion of RMIT-O17 and ΔluxS-O17 to that were adherent to INT 407 (Figure 4.15). The mean % adhesion value of RMIT-O17 at MOI 100 was 0.13 while the mean % adhesion value of ΔluxS-O17 it was 0.087. Interestingly, the mutant exhibited significantly less invasion that the wild type (*p* < 0.02). The invasive index of RMIT-O17 was 0.281 while for ΔluxS-O17 was zero. According to Larson *et al.* (2008), a *C. concisus* strain that has an invasive index ≥1 was defined as an enteric-invasive *C. concisus* (EICC) strain. The wild type RMIT-O17 was detected as non-EICC, but the invasive index of RMIT-O17 was significant (*p* < 0.02) compared to the mutant ΔluxS-O17 by t-test.
Figure 4.15: Adherence of RMIT-O17 and ΔluxS-O17 to INT 407 epithelial cells showing no statistically significant difference. Adherence is expressed as the percentage of the inoculum that remained associated with the epithelial cells. Red bar represents RMIT-O17, striped red bar represents the ΔluxS-O17 mutant. Experiments were performed three times in triplicate, and error bars represent standard error of the mean.
**Figure 4.16:** Invasion of RMIT-O17 and ΔluxS-O17 to INT 407 epithelial cells showing the mutant exhibited significantly less invasion than the wild type ($p < 0.02$). Invasion is expressed as the percentage of the inoculum that invaded in the epithelial cells and showing significant difference between the wild type and mutant. Red bar represents RMIT-O17. Experiments were performed three times in triplicate, and error bars represent standard error of the mean.
**Figure 4.17**: Invasive index of RMIT-O17 and ΔluxS-O17 to INT 407 epithelial cells. Invasion index is expressed as the percentage of the inoculum that invaded in the epithelial cells divided by the cells adhered and showing significant difference between the wild type and mutant. Red bar represents RMIT-O17, striped red bar represents the ΔluxS-O17 mutant. Experiments were performed three times in triplicate, and error bars represent standard error of the mean.
4.6.3.4 Confocal laser scanning microscope (CLSM)

RMIT-O17 and ΔluxS-O17 were grown on cover slip in 6 well plates in Brucella broth for 4 days in hydrogen-enriched condition as mentioned in Section 2.3.2.

The biofilm of RMIT-O17 in day 4 was a complex of live and few dead cells present within small clumps, while no similar structures were found in the ΔluxS-O17 biofilm (Figure 4.18). All of the cells of ΔluxS-O17 attached to the coverslip as single cells and only live cells were seen in the biofilm.

4.6.3.5 Scanning electron microscope (SEM)

RMIT-O17 and ΔluxS-O17 were grown on 10 mm round shaped cover slip in 12 well plates in Brucella broth for 4 days in hydrogen-enriched condition as mentioned in Section 2.3.2.

The biofilm of the wild type RMIT-O17 in day 4 was a clump of C. concisus cells present within aggregations attached to each other, while no similar structures were found in the ΔluxS-O17 biofilm (Figure 4.19). All of the cells of ΔluxS-O17 attached to the coverslip mostly as single cells and thereof no biofilm complex was formed by the mutant ΔluxS-O17, which indicates the quorum-sensing is not occurring in the ΔluxS mutant C. concisus strain.
Figure 4.18: Biofilm of RMIT-O17 (A) and ΔluxS-O17 (B) by CLSM at 200x magnification and 3.4 zooming showing less biofilm production by luxS-mutant compared to the wild type.

Figure 4.19: Biofilm of RMIT-O17 (A) and ΔluxS-O17 (B) by SEM at 5000x magnification showing less biofilm production by luxS-mutant compared to the wild type.
4.7 Discussion

This study provides valuable information on the role of the luxS gene in C. concisus regarding biofilm formation and other virulence-related characteristics. A luxS knock-out mutant was generated in a biofilm-forming strain of C. concisus. This is the first study investigating the role of luxS in biofilm formation in C. concisus.

RMIT-O17, the highest biofilm producer C. concisus in the collection was selected to knock out luxS gene to investigate its role in biofilm formation in C. concisus. As luxS is a relatively small gene (516 bp), primers were designed from the flanking region of this gene. C. concisus is a genetically diversified organism and gene shuffling is a common phenomenon, which was shown by Deshpande et al. (2013) where six whole genomes of C. concisus were sequenced and compared with the reference strain C. concisus 13826. The alignment of the available C. concisus genomes (Table 4.2) showed that the presence of two different downstream genes after luxS was genospecies specific, with a gene encoding for a putative protein in genospecies A strains or a gene encoding for a transduction protein in genospecies B strains. However it is worthy to mention that the genome of UNSWCS, a genospecies B strain, followed the gene arrangement of genospecies A, which provided further evidence on the divergence of this bacterium. Therefore when primers were designed from the flanking regions, two reverse primers were selected for RMIT-O17 luxS-flanking region. However, RMIT-O17 was found to be consistent with other genospecies A strains and amplified a PCR product with the reverse primer designed from the putative protein.

The knock-out of luxS was performed successfully in RMIT-O17 and the mutant was tested for biofilm formation, and virulence factors as invasion and adhesion, motility
Chapter 4: Knock-out of luxS in C. concisus

and phenotypic characteristics of biofilm by CLSM and SEM. Biofilm formation by the ΔluxS-O17 strain was found to be significantly reduced ($p < 0.05$) (Figure 4.13), which supports a role for luxS in biofilm formation in C. concisus. The reduction of biofilm was also supported by the CLSM and scanning electron micrograph (SEM) (Figure 4.18 and 4.19). An association between AI-2 signalling and biofilm formation has also been reported in Streptococcus gordonii (McNab et al., 2003), S. mutans (Merritt et al., 2003), (Yoshida et al., 2005) Streptococcus anginosus (Petersen et al., 2006), Klebsiella pneumoniae (Balestrino et al., 2005) and close related species H. pylori (Cole et al., 2004). Previously, Merritt et al. (2003) had found luxS mutant to be defective in AI-2 signal, hence quorum-sensing and biofilm produced by the luxS mutant were different from the wild type S. mutans. Yoshida et al. (2005) found that biofilm formation by the luxS mutant was markedly decreased compared to the wild type S. mutans. They also showed that biofilms of the luxS mutant formed larger clumps in defined medium supplemented with 0.5% sucrose compared to the parental strain by scanning electron microscopy, which is in agreement with our result for C. concisus.

Studies on C. jejuni biofilms by Reeser et al. (2007) found both flagella and quorum-sensing are required for maximum biofilm formation, as both C. jejuni flaAB and luxS mutants were significantly reduced in their ability to form biofilms compared to the wild type strain. In contrast Lavrencic et al. (2012) did not report a correlation between motility and biofilm formation in C. concisus. In this study the phenotypic properties of eight biofilm-forming C. concisus strains was observed but no molecular investigation was done. The C. concisus strains used in this study were shown to be able to produce biofilm to at least some degree regardless of their source of isolation. But the motility was found to be strain-specific and it was suggested that strains with higher motility have a greater
chance to swim through the intestinal mucus layer and reach the epithelial surface and cause diseases.

It was also found ΔluxS-O17 became less motile in semi-solid agar compared to the wild type RMIT-O17 (Figure 4.14). In *Vibrio alginolyticus* LuxS was found to regulate flagella biogenesis (Tian *et al.*, 2008) thus also regulating motility in *V. alginolyticus*. In *Vibrio harveyi*, motility assay and analysis of gene expression indicated the involvement of the quorum-sensing system and autoinducer synthase mutants (ΔluxO) showed significantly lower swimming motility and expression of flagellar genes than the wild type, which was restored by adding synthetic signalling molecules (Yang and Defoirdt, 2015). In *C. jejuni* inactivation of LuxS also exhibited decreased motility in semisolid media, suggesting a role for quorum-sensing in the regulation of motility (Elvers and Park, 2002). The authors suggested that this system serves as a global regulatory mechanism for basic physiological function and virulence factors. Consistent with previous studies, Holmes *et al.* (2009) reported that *C. jejuni* luxS mutant has comparable growth rates to the parental strain and exhibited decreased motility halos in both MEM-alpha and MHB. Very recently, Adler *et al.* (2014) reported that different phenotypes of *C. jejuni* ΔluxS mutants depend on strain background, mutation strategy and culture conditions. This group also complemented luxS with synthetic AI-2 or homocysteine as well as the combination of both and genetic complementation and partially restored swarming abilities of *C. jejuni* 81-176ΔluxS.

The ΔluxS-O17 was non-invasive in the INT 407 intestinal cell line (Figure 4.16) although there was no significant difference in adhesion between the mutant and wild type. This is a very interesting finding as luxS is a signalling molecule and it has a role in
metabolism for *C. concisus*. Similar results were found in *C. jejuni* where *C. jejuni ΔluxS* mutant has been shown comparable growth rate, resistance to oxidative stress and ability to invade Caco-2 cell monolayers to the parental strain (Elvers and Park, 2002). However, the authors could not find any significant difference between the wild type and mutant regarding invasion in Caco-2 monolayers. On the other hand, Quinones *et al.* (2009) found significantly reduced colonization of the chick lower gastrointestinal tract, chemotaxis toward organic acids, and *in vitro* adherence to LMH chicken hepatoma cells after inactivation of luxS in *C. jejuni* strain 81-176. It was suggested that AI-2 production in *C. jejuni* contributes to host colonization and interactions with epithelial cells. LuxS plays a role in *E. coli* O157:H7 in attaching and effacing lesions in pigs (Jordan *et al.*, 2005). Inactivation of quorum-sensing regulator in *Vibrio vulnificus*, a food-borne pathogen, caused reduced virulence and colonization capacity in the infection of mice. Moreover, the mutant exhibited significantly reduced biofilm detachment to INT 407 host cells (Kim *et al.*, 2013).

In this study, INT 407 cell line was used for comparative invasion and adhesion assay for ΔluxS mutant and the wild type. It is worthy to note that although it has been indicated that the INT 407 cell line contains human papilloma virus (HPV-18) ([http://www.atcc.org/products/all/CCL-6.aspx#characteristics](http://www.atcc.org/products/all/CCL-6.aspx#characteristics)), it did not affect our results on the invasiveness of *C. concisus*.

To summarise, the luxS gene was deleted in *C. concisus* and led to significant reductions in biofilm formation, motility and invasion in INT 407. In order to fulfill molecular Koch’s postulates (Falkow, 2004), it is necessary to complement the gene in the mutant to retrieve the function of the gene either by genetic or chemical complementation.
The complementation of the gene could provide the evidence that this gene is solely responsible for biofilm formation in *C. concisus*. Successful genetic complementation of *luxS* was shown by Quinones *et al.* (2009) and they could restore the adherence property of *C. jejuni*. Later, Adler *et al.* (2014) restored the swarming property of *C. jejuni* partially by complementation. An attempt was taken to do so, but unfortunately for time constraints in a PhD project, the experiment could not be finished as the construction of an *E. coli*-*C. concisus* shuttle vector is a complicated and laborious experiment. This work remains a priority and needs to be done in near future.
Chapter Five

Comparative Genomics of selected oral and intestinal Campylobacter concisus strains

5.1 Introduction

C. concisus is genetically diverse microorganism. Since the discovery of this microorganism in 1981 Tanner et al. (1981), there have been many attempts to classify C. concisus using different approaches, for example, DNA: DNA hybridization (Vandamme et al., 1989); RAPD (Randomly amplified polymorphic DNA analysis) (Van Etterijck et al., 1996; Matsheka et al., 2006); AFLP analysis (On and Harrington, 2000); lectin typing (Aabenhus et al., 2002a); and pulsed field gel electrophoresis-based genotyping (Matsheka et al., 2002). Vandamme et al. (1989) reported that faecal C. concisus isolates exhibited only 42–50% DNA: DNA hybridization values with strains of oral origin. Until 2011, there was only one genome sequence available for C. concisus 13826, a strain isolated from faeces of acute gastroenteritis patient and the genome was fully sequenced in 2007. The second C. concisus strain (UNSWCD) isolated from intestinal biopsy of a patient with Crohn's disease was sequenced (Deshpande et al., 2011) and only 76% of genes were homologues between C. concisus 13826 and UNSWCD (Kaakoush et al., 2011a). The authors of this study have concluded that C. concisus 13826 is genetically atypical to other C. concisus strains (Kaakoush et al., 2011a). Later this group sequenced six more C. concisus strains isolated from various clinical sources (two CD, one chronic gastroenteritis, two acute gastroenteritis patients, one from a healthy control) and showed evidence of gene shuffling in C. concisus (Deshpande et al., 2013).
Chapter 5: Comparative genomics of oral and intestinal *C. concisus*

There is evidence about the role of *C. concisus* as an intestinal pathogen (Vandamme *et al.*, 1989; Van Etterijck *et al.*, 1996; Lastovica and le Roux, 2000; Nielsen *et al.*, 2013b) and also evidence of the presence of this microorganism in the oral cavity of healthy individuals (Tanner *et al.*, 1981; Kamma *et al.*, 2000b; Kamma *et al.*, 2000a; Zhang *et al.*, 2010). In Chapter 3, it was shown that *C. concisus* strains isolated from the oral cavity produced more biofilm than intestinal isolates, although the *luxS* gene potentially responsible for biofilm formation was found in all *C. concisus* analysed sequences. This has raised the argument that the diversity of oral and intestinal *C. concisus* genome sequences needs to be thoroughly investigated. Currently, there are eight genome sequences of intestinal *C. concisus* and only one oral *C. concisus* strain (isolated from gingival sulcus) available in the NCBI database. However, the genome of *C. concisus* isolates from the oral cavity of a healthy individual or oral isolates from Crohn’s disease patients have not yet been sequenced. To further examine the importance of *C. concisus* heterogeneity with respect to biofilm formation, disease potential and source of isolation, the genomes of four selected *C. concisus* strains were sequenced and comparative analyses of these four *C. concisus* and two reference sequences were performed, which allowed comparison of strains isolated from the oral cavity and gastrointestinal tract.

The aims of the experimental procedure performed in this chapter were to:

1. Perform whole genome shotgun sequencing of selected oral and intestinal *C. concisus* strains and comparative genomic analysis to examine the importance of *C. concisus* heterogeneity with respect to source of isolation,

2. Define the pan and core genomes of *C. concisus*,

3. Detect, annotate and map plasmids in the sequenced genomes.
5.2 Materials and Methods

5.2.1 C. concisus strain selection for WGS

Four C. concisus strains were sequenced in this study. They were isolated from patients with IBD (AUS22-Bd2), gastroenteritis (RCH 26), Crohn’s disease (RMIT-JF1) and a healthy individual (RMIT-O17). Ethics approvals for the strains used in this chapter are included in Section 2.2.6. Written consent was obtained from all subjects, or their guardians, who were participating in the study.

5.2.2 DNA extraction for WGS

C. concisus strains were grown on HBA in microaerophilic hydrogen-enriched environment as explained in Section 2.2.7. Bacterial genomic DNA (gDNA) was extracted using a Wizard Genomic DNA Purification Kit (Promega BioScience, CA, USA) following the manufacturer’s instructions and stored at -20°C in 10 mM Tris buffer. Extracted DNA was electrophoresed in agarose gel to observe the quality of the extraction and presence of plasmids. DNA concentration was measured using Qubit (Invitrogen by Life Technologies, Singapore) according to the manufacturer’s instructions (Section 2.6.1.1). Based on the Qubit reading, 50-100 ng of gDNA were used to prepare the library for sequencing purposes.

5.2.3 Whole genome sequencing (WGS)

5.2.3.1 WGS using Ion Proton platform

Genome sequencing of RCH 26 and RMIT-JF1 were performed using an Ion Proton next generation sequencer (Thermo fisher Scientific Inc.) at RMIT University, Melbourne. The libraries were prepared using an Ion PI Sequencing 200 Kit v3 (Life Technologies) following the manufacturer’s instructions (Section 2.6.2.1).
preparations were qualitatively and quantitatively evaluated using a Bioanalyzer (Agilent® 2100, Germany) before being used in template preparation steps (Section 2.6.2.2). The High Sensitivity DNA Analysis Kit (Agilent Technologies, Germany) was used for sample and chip preparations according to the manufacturer’s instructions. The sequencing template preparation for *C. concisus* strains was performed using the Ion OneTouch™ 2 system and Ion PI™ Template OT2 200 Kit v3 (Life Technologies, USA) according to the manufacturer’s instructions (Ion Torrent, 2014).

### 5.2.3.2 **WGS using Illumina platform**

Genome sequencing of RMIT-O17 and AUS22-Bd2 were performed using an Illumina MiSeq sequencer at RMIT University, Melbourne. The libraries were prepared using the Nextera XT DNA library preparation kit (Illumina, 2012) following the manufacturer’s instructions described in Section 2.6.3.1. One MiSeq® Reagent Kit v3 (2x300 bp paired-end reads) was used to perform the run on the Illumina MiSeq.

### 5.2.4 **Draft genome assembly**

#### 5.2.4.1 **De novo assembly of RMIT-JF1**

The *de novo* assembly for RMIT-JF1 was performed by AGRF, Melbourne. The fastq file from the sequencer was sent to AGRF where it was screened using the Fastqc tools ([http://www.bioinformatics.babraham.ac.uk/projects/fastqc/](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)). Ion torrent barcodes and adaptor were clipped using ea-utils fastq-mcf (FASTX tool kit: [http://hannonlab.cshl.edu/fastx_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)). The reads were also quality trimmed at ends using fastq_quality_trimmer (FASTX tool kit: [http://hannonlab.cshl.edu/fastx_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)) at quality cut-off of Q20. Cleaned reads were used for assembling contigs for the sample. The *de novo* assembly was performed using the open source software package Mira4.
Chapter 5: Comparative genomics of oral and intestinal *C. concisus*

(http://www.chevreux.org/projects_mira.html). An in house developed script was used to split sequences into short reads < 120bp and long reads ≥ 120 and perform a hybrid assembly using Mira4 using solexa settings for short reads and iontorrent settings for long reads.

5.2.4.2 *De novo* assembly of RMIT-O17 and RCH 26

The *de novo* assembly for RCH 26 and RMIT-O17 were performed using CLC Genome Workbench 8.5.1. In CLC Genomics Workbench 8.5.1, the raw FASTQ reads were first trimmed to quality score limit 0.001 (Q30) with maximum 2 ambiguous nucleotides and reads with length below 15 nucleotides were discarded. These trimmed reads were then *de novo* assembled with optimized bubble size and word size (*k*-mer) based on assembly output parameters such as high N50, low total number of contigs, high average contig length, and high percentage of mapped reads. The set up was kept in mapping mode ‘map reads back to contigs’ with scaffolding and a minimum contig length of 200 nucleotides.

5.2.4.3 *De novo* assembly of AUS22-Bd2

The *de novo* assembly for AUS22-Bd2 was performed using A5-miseq pipeline which included adapter trimming, quality filtering, error correction, contig and scaffold generation, and detection of misassemblies (Coil *et al.*, 2015).

5.2.4.4 GenBank submission

The whole genome shotgun projects of RMIT-JF1, RMIT-O17, RCH 26 and AUS22-Bd2 have been deposited at NCBI GenBank. The GenBank files for individual genome sequences were extracted for annotation and analysis.
5.2.5 Pairwise comparison of C. concisus strains and analysis

Genome scale alignments were performed by pairwise comparison of the four C. concisus strains with the reference strains C. concisus 13826 and ATCC 33237\textsuperscript{T} using the Mauve alignment tool (Darling et al., 2004). Mauve is a software package that attempts to align orthologous and xenologous regions among two or more genome sequences that have undergone both local and large-scale changes. Inside each block Mauve draws a similarity profile of the genome sequence. The height of the similarity profile corresponds to the average level of conservation in that region of the genome sequence. Areas that are completely white are not aligned and probably contain sequence elements specific to a particular genome. The height of the similarity profile is calculated to be inversely proportional to the average alignment column entropy over a region of the alignment.

5.2.6 Annotation of the assembled C. concisus strains and analysis

The individual GenBank files of four C. concisus genomes generated by NCBI GenBank were extracted and used for annotation. The GenBank files were submitted to the Rapid Annotation using Subsystem Technology (RAST) (Aziz et al., 2008), a service for annotating bacterial and archaeal genomes, in relation to gene definition and annotation for individual assemblies. Additionally, some genes of interest were confirmed by homology BLAST searches of the NCBI database. The protein sequences extracted from RAST were further blasted, mapped and annotated using the annotation and analysis tool Blast2GO (Conesa and Gotz, 2008).

A comparative sequence based analysis of the four C. concisus strains was performed and a circular view was generated using the Seed Viewer (Overbeek et al., 2014) compared with reference sequences of C. concisus 13826 and ATCC 33237\textsuperscript{T}. Seed
viewer uses the annotation stored in the RAST server for the comparison. Orthologous gene search was performed using the genes from the previously identified two plasmids within *C. concisus* 13826 to check for conservation in the sequenced *C. concisus* genomes using OrthoVenn (Wang *et al.*, 2015) (http://aegilops.wheat.ucdavis.edu/OrthoVenn/).

### 5.2.7 Plasmid detection, mapping and construction

Contigs smaller than 40 kb in size in *de novo* assemblies of the four *C. concisus* sequences were checked by a web based tool ‘Tandem Repeat Finder’ (Benson, 1999) for the origin of replication (*ori*) within the contig. An origin of replication is a sequence of DNA at which replication is initiated on a chromosome, plasmid or virus. Any contigs containing tandem repeats were identified and a pair of inverse primers was designed from the edge of the contigs to check whether the inverse PCR primers could amplify a PCR product. If any amplification was successful, the PCR products were purified and sequenced by Sanger sequencing to complete the circular structure of the plasmid. Inverse PCR primers used in this study are presented in Table 5.1. The newly constructed plasmids were characterized by annotation, and mapping and the circular maps were viewed using CGview Server (http://stothard.afns.ualberta.ca/cgview_server/) (Stothard and Wishart, 2005).

### 5.2.8 Pan- and core-genome analysis

The core and pan genomes of four *C. concisus* strains (AUS22-Bd2, RMIT-O17, RMIT-JF1 and RCH 26) were defined using BPGA 1.2 (Bacterial Pan Genome Analysis Pipeline), a perl based pipeline (Chaudhari *et al.*, 2016). The reference strains *C. concisus* 13826 and ATCC 33237<sup>T</sup> were included in the analysis. USEARCH clustering tool was used to obtain pan-genome profiles of bacterial gene pools obtained from RAST. BPGA
uses MUSCLE for alignments and tree generation. Another prerequisite gnuplot 4.6.6 which is required for plotting the graph was installed manually on the system. Protein sequences extracted from the RAST annotation for four of the *C. concisus* strains (AUS22-Bd2, RMIT-O17, RMIT-JF1 and RCH 26) and protein sequences of two reference strains from the NCBI database were used for analysis. The analysis was performed using the default parameters:

- Clustering: USEARCH (Identity cut off = 50%)
- No. of combinations: 30
- Atypical GC Content Analysis: $2 \times \delta$ (Standard Deviation)
- Type of phylogeny tree: Neighbor Joining Tree (NJ).

The analysis gave representative sequences for core (genes present in all strains), accessory (genes present in two or more but not all strains) and unique gene sets (genes present only in one strain) and these gene sets were added to get the final pan genome set for these six *C. concisus* strains.

### 5.2.9 Enrichment analysis using blast2GO

Gene ontology terms (GO terms) were assigned to ORFs identified by RAST from Section 5.2.6 using the annotation and analysis tool Blast2GO, Version 3.3 (Conesa and Gotz, 2008). Blast2GO was also used for functional enrichment analyses of *C. concisus* core and pan genomes (derived by using BPGA 1.2 Pipeline in Section 5.2.7). Following annotation, the statistical analysis package in Blast2GO which uses Fisher’s exact test using a $p$-value cut-off of 0.05 with multiple testing correction of false discovery rate (FDR) (Al-Shahrour *et al*., 2004) was applied and enrichment graphs were generated.
Table 5.1: Primer sequences used for inverse PCR of plasmids in this chapter.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Size</th>
<th>Primer pair</th>
<th>Sequence</th>
<th>Template</th>
<th>T&lt;sub&gt;A&lt;/sub&gt; (°C)</th>
<th>Target region</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inv con17R</td>
<td>18</td>
<td>1</td>
<td>TTCCG GTGGGCATATCAG</td>
<td>RCH 26 gDNA</td>
<td>55</td>
<td>Contig 17</td>
<td>This study</td>
</tr>
<tr>
<td>Inv con17F</td>
<td>18</td>
<td></td>
<td>CAGCTTGGAAGCTAAGAG</td>
<td>gDNA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inv con20R</td>
<td>22</td>
<td>2</td>
<td>GGC GTAGCCTTCTATCTCATC</td>
<td>RCH 26 gDNA</td>
<td>56</td>
<td>Contig 20</td>
<td>This study</td>
</tr>
<tr>
<td>Inv con20F</td>
<td>22</td>
<td></td>
<td>CGGA AAGGC GG TTGATTGA</td>
<td>gDNA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.2.10 Detection of clusters of orthologous gene (COG) in four *C. concisus* strains

Orthologous genes are clusters of genes in different species originated by vertical descent from a single gene in the last common ancestor. The different and overlapping orthologous clusters among four *C. concisus* strains were analysed and displayed in a Venn diagram using a web platform named OrthoVenn ([http://probes.pw.usda.gov/OrthoVenn](http://probes.pw.usda.gov/OrthoVenn)) (Wang et al., 2015). It provided circles representing each strain with overlapping regions that illustrate the genes or gene clusters that are unique to, or shared between each strain. It uses UBLAST (v7.0.1090) to do the all-against-all similarity search and orthAgogue (v1.0.3) for identification of putative orthology and inparalogy relations. The intersection of orthologous clusters were analysed by GO Slim annotation and UniProt search. The alignment of the protein sequences in each cluster was performed by using the multiple sequence alignment tool MUSCLE.

The analysis was performed using the following parameters:

- E-value: $1 \times 10^{-10}$
- Inflation value: 1.5

5.2.11 Genes/proteins associated with virulence

From a review of the *C. concisus* literature, it was established that there are some toxins and virulence related genes in specific strains, including Zot (Zonula Occludens Toxin and tight junction toxin), secreted and cell bound hemolysins like OMPLA (outer membrane phospholipase A) encoded by the *pldA* gene, S-layer RTX or NhaC (protein related to adhesion), genes responsible for motility, such as *flaA, flaB* and *flaC*, exotoxin 9 (related to invasion) and putative virulence genes *cjaC, cjaA* (both are immunogenic surface protein) and *dnaJ* (encodes a stress protein) An investigation of the presence of
these virulence related genes or toxins were performed in the genomes of AUS22-Bd2, RCH 26, RMIT-JF1 and RMIT-O17.

5.3 Results

5.3.1 Confirmation of C. concisus DNA quality for genome sequencing

For whole genome sequencing, the quality of the extracted genomic DNA was confirmed by visualising the sample as shown in Figure 5.1. In AUS22-Bd2, RMIT-JF1 and RMIT-O17 only one chromosomal band was observed, while in RCH 26 two additional bands similar to plasmids were observed along with the chromosomal band in the agarose gel. However, the size of the plasmid/s could not be confirmed due to their supercoiled structure.
Figure 5.1: Gel electrophoresis of gDNA extracted from the four *C. concisus* strains used for whole genome sequencing showing the presence of plasmids only in RCH 26. Lane 1: RCH 26; Lane 2: AUS22-Bd2; Lane 3: *Pst*I digested λ DNA ladder; Lane 4: RMIT-JF1; Lane 5: RMIT-O17
5.3.2 Whole genome sequencing

5.3.2.1 RMIT-JF1 and RCH 26

The preparation of RCH 26 and RMIT-JF1 genomic DNA libraries were carried out using the Ion Xpress™ Plus gDNA Fragment Library Preparation kit V2 (Section 2.6.3.1). The fragmentation of the gDNA and ligation of the barcoded adapters was followed by size selection of the DNA fragments to optimise the read length of the sequencing, the size selection targeted fragments with sizes of 200-230 bp (150 bp library excluding barcodes and adapters). The prepared libraries of RCH 26 and RMIT-JF1 were analysed on Agilent 2100 Bioanalyzer to determine the molar concentration of the DNA samples and to assess the size distribution of the libraries. The DNA molar concentrations (Table 5.2) were used to calculate the dilution required for each DNA sample to be used in the template preparation. The Bioanalyzer electropherograms are presented in Figures 5.2(A & B). The peak at 35 bp is the lower marker and the 10,380 bp peak represents the upper marker. The major peaks for fragments with a size range of approximately 200 to 300 bp were well represented in the C. concisus strain libraries. A summary of the sequencing data produced by Ion proton for each strain is presented in Table 5.3.

5.3.2.2 RMIT-O17 and AUS22-Bd2

The library was prepared using a Nextera XT DNA library preparation kit (Illumina, 2012) and sequencing of RMIT-O17 and AUS22-Bd2 was performed in Illumina MiSeq. Sequencing data of these two strains are listed in Table 5.4.
Table 5.2: Quantification of RMIT-JF1 and RCH 26 prepared libraries by High Sensitivity DNA Assay using Agilent 2100 Bioanalyzer; Software: 2100.

<table>
<thead>
<tr>
<th>C. concisus DNA library</th>
<th>Size (bp)</th>
<th>Concentration (pg/μl)</th>
<th>Molarity (pmol/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMIT-JF1</td>
<td>260</td>
<td>81.52</td>
<td>474.6</td>
</tr>
<tr>
<td>RCH 26</td>
<td>256</td>
<td>93.31</td>
<td>552.4</td>
</tr>
</tbody>
</table>

Table 5.3: Summary of the sequencing data of RMIT-JF1 and RCH 26 reported from the Ion Proton sequencing platform.

<table>
<thead>
<tr>
<th>Barcode Name</th>
<th>Sample</th>
<th>Bases</th>
<th>≥Q20</th>
<th>Reads</th>
<th>Mean Read Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>IonXpress 001</td>
<td>RMIT-JF1</td>
<td>1,138,778,761</td>
<td>887,751,850</td>
<td>8,439,170</td>
<td>134 bp</td>
</tr>
<tr>
<td>IonXpress 002</td>
<td>RCH 26</td>
<td>1,046,037,810</td>
<td>801,567,161</td>
<td>7,345,932</td>
<td>142 bp</td>
</tr>
</tbody>
</table>
Figure 5.2: Electropherogram summary of (A) RMIT-JF1; (B) RCH 26 genomic DNA libraries. The peak at 35 bp is the lower marker and the 10,380 bp peak represents the upper marker. The major peaks for fragments with a size of approximately 250 bp were well represented in *C. concisus* genomic DNA libraries.
5.3.3 Genome assembly of C. concisus strains

*De novo* assembly of RMIT-JF1 was performed by AGRF and 25 contigs were generated with a genome size of 1.94 Mbp genome. Genomic read data of RCH 26 from the Ion proton and RMIT-O17 were assembled using CLC genomic workbench and 24 and 34 contigs were obtained with 1.91 and 1.84 Mbp genome, respectively. The best assembly for RCH 26 was obtained with word size 26 and bubble size 144, while for RMIT-O17 the word size was 17 and bubble size was 98. However, 2 contigs in RCH 26 assembly were found to be located within a plasmid (Section 5.3.6.2); therefore, final assembly of RCH 26 contained 23 contigs. *De novo* assembly of AUS22-Bd2 was kindly performed by Professor Rob Moore, RMIT University, using A5-miseq pipeline and 42 contigs were obtained with a size of 1.83 Mbp genome. All data obtained from assemblies from four *C. concisus* strains and the accession numbers are shown in Table 5.4.

The assembled genomes of four *C. concisus* strains were compared pairwise with the reference strain *C. concisus* 13826 (Figure 5.3) and ATCC 33237^T^ (Figure 5.4) using the Mauve alignment tool (Darling et al., 2004). This global alignment was performed to investigate any large-scale changes during evolution and to show paired synteny alignments, which indicate the genome shuffling of *C. concisus* strains. This analysis revealed varying degrees of genome shuffling, shown by line connections of locally collinear blocks (LCB) across genome pairs, and inversions marked below the reference axis. When the four *C. concisus* genomes sequenced in this study were compared with *C. concisus* 13826 very high level of gene shuffling was observed as indicated by the connection lines. The LCB were smaller in size and a greater number of inverted LCBs were found in these comparisons (Figure 5.3).
Table 5.4: Information on whole genome sequences of AUS22-Bd2, RMIT-O17, RCH 26 and RMIT JF1.

<table>
<thead>
<tr>
<th>Strain</th>
<th>AUS22-Bd2</th>
<th>RMIT-O17</th>
<th>RCH 26</th>
<th>RMIT-JF1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Disease</strong></td>
<td>IBD</td>
<td>Healthy</td>
<td>Gastroenteritis</td>
<td>Crohn’s disease</td>
</tr>
<tr>
<td><strong>Isolation site</strong></td>
<td>Duodenum</td>
<td>Oral cavity</td>
<td>Stool</td>
<td>Oral cavity</td>
</tr>
<tr>
<td><strong>Source</strong></td>
<td>Austin Hospital, Melbourne</td>
<td>RMIT University, Melbourne</td>
<td>The Royal Children’s Hospital, Melbourne</td>
<td>RMIT University, Melbourne</td>
</tr>
<tr>
<td><strong>Sequencer</strong></td>
<td>Illumina MiSeq</td>
<td>Illumina MiSeq</td>
<td>Ion Proton</td>
<td>Ion Proton</td>
</tr>
<tr>
<td><strong>Bases ≥ Q40</strong></td>
<td>1,835,029</td>
<td>1,839,227</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td><strong>De novo assembly</strong></td>
<td>A5-miseq pipeline</td>
<td>CLC Genomic Workbench</td>
<td>CLC Genomic Workbench</td>
<td>Mira 4</td>
</tr>
<tr>
<td><strong>Reads x2</strong></td>
<td>4,539,778</td>
<td>1,867,530</td>
<td>4,803,072</td>
<td>3,792,122</td>
</tr>
<tr>
<td><strong>Contigs</strong></td>
<td>42</td>
<td>34</td>
<td>23</td>
<td>25</td>
</tr>
<tr>
<td><strong>N50</strong></td>
<td>125,724</td>
<td>486,729</td>
<td>273,604</td>
<td>338,852</td>
</tr>
<tr>
<td><strong>Scaffolds</strong></td>
<td>42</td>
<td>34</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td><strong>Coverage</strong></td>
<td>577</td>
<td>341</td>
<td>555</td>
<td>350</td>
</tr>
<tr>
<td><strong>Genome size (Mbp)</strong></td>
<td>1.83</td>
<td>1.84</td>
<td>1.91</td>
<td>1.94</td>
</tr>
<tr>
<td><strong>No. of coding sequences</strong></td>
<td>1,859</td>
<td>1,902</td>
<td>2,141</td>
<td>2,048</td>
</tr>
<tr>
<td><strong>% GC content</strong></td>
<td>37.5</td>
<td>37.6</td>
<td>37.5</td>
<td>37.62</td>
</tr>
<tr>
<td><strong>Accession no</strong></td>
<td>LVXF00000000</td>
<td>LVLC00000000</td>
<td>LVWL00000000</td>
<td>JXUP00000000</td>
</tr>
</tbody>
</table>

NA: Not applicable
Figure 5.3: Pairwise comparisons of RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17 with *C. concisus* 13826 using the Mauve alignment tool.
Figure 5.4: Pairwise comparisons of RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17 with *C. concisus* ATCC 33237<sup>T</sup> using the Mauve alignment tool.
When these four *C. concisus* genome sequences were compared with ATCC 33237<sup>T</sup> (Figure 5.4), less genome shuffling was observed. LCBs were relatively larger and no inversion was found when ATCC 33237<sup>T</sup> was compared to RHC 26, RMIT-O17 and AUS22-Bd2. A high level of similarity and contiguity between ATCC 33237<sup>T</sup> and AUS22-Bd2 was observed (Figure 5.4).

### 5.3.4 Annotation of the assembled *Campylobacter concisus* strains and analysis

Rapid Annotation using Subsystem Technology (RAST) (Aziz *et al*., 2008) was used for gene definition and annotation for AUS22-Bd2, RMIT-O17, RMIT-JF1 and RCH assemblies. Total number of genes, proteins, %GC content and the genome size obtained are presented in Table 5.4. The annotated genomes were viewed and analysed in Seed Viewer server (Overbeek *et al*., 2014). The "subsystem" in RAST is a collection of functional roles that make up a metabolic pathway, a multi-subunit complex (e.g., the ribosome), a specific class of proteins (e.g., signal transduction), or any other set of functional roles that the annotator believes are related in some biologically meaningful way (Overbeek *et al*., 2014). Subsystem category distribution and subsystem feature counts in the genome of AUS22-Bd2, RMIT-O17, RMIT-JF1 and RCH 26 is presented in the pie chart in Figure 5.5 (A, B, C and D). The number of proteins present in each subsystem feature is presented within bracket. In the bar graph of Subsystem coverage the green colour indicates presence in the subsystem, and the blue colour indicates absence in the subsystem. Among the features, the count of cell wall and capsule, nucleosides and nucleotides, protein metabolism, amino acids and derivatives were higher in RCH 26 compared to other three genomes. However, in RCH 26 there was no feature count for secondary metabolism, while the other three genomes had four of those in their sequences. AUS22-Bd2 had higher counts of DNA metabolism compared to other genome sequences.
Chapter 5: Comparative genomics of oral and intestinal *C. concisus*
Figure 5.5: Subsystem statistics and feature counts of AUS22-Bd2 (A), RMIT-JF1 (B), RMIT-O17 (C) and RCH 26 (D) genomes obtained from the Seed Viewer (Overbeek et al., 2014).
A genome-based comparison was performed on the annotated genomes of four *C. concisus* strains with two reference sequences for this species, *C. concisus* 13826 and ATCC 33237T and summarised in Figure 5.6 (A & B). RAST includes a completely annotated genome for the reference *C. concisus* 13826. ATCC 33237T genome sequence was manually annotated using RAST to be used for this comparison. The comparison is presented in a circular view generated by the Seed viewer. The sequence from the outer ring to inner ring is RMIT-JF1, AUS22-Bd2, RMIT-O17 and RCH 26 compared with *C. concisus* 13826 (Figure 5.6 A) and ATCC 33237T (Figure 5.6 B). The identity percentage of the protein sequences are presented in colour from purple (100%) to red (10%) compared with the reference sequences. Absence of any gene in the ring is indicated by a white colour. From the colour intensity of Figure 5.6 (A & B), it is observed that RMIT JF1, AUS22-Bd2, RMIT-O17 and RCH 26 are more close to ATCC 33237T, which was also observed in the pairwise alignment by Mauve.

### 5.3.5 Investigation of *C. concisus* 13826 plasmid genes in the studied *C. concisus* strains

Two plasmids (pCCON31, n = 33 and pCCON16, n = 23) present in the reference strain *C. concisus* 13826 were checked for conservation and positioning across the four strains in this study. The number of genes found in each strain is presented in Table 5.5. Both AUS22-Bd2 and RMIT-JF1 had 3 and 6 orthologous genes with pCCON31 and pCCON16 respectively. Three orthologous hits from pCCON31 are a hypothetical protein (EAT97541.1), para protein (EAT97550.1) and site-specific recombinase, phage integrase family (EAT97534.1).
Chapter 5: Comparative genomics of oral and intestinal \textit{C. concisus}

\textbf{Figure 5.6:} A sequence based comparison of RMIT JF1, AUS22-Bd2, O17 and RCH 26 annotated genomes with \textit{C. concisus} ATCC 33237$^\text{T}$ (A) and \textit{C. concisus} 13826 (B). Percent protein identity is showing by colour at the bottom.
Table 5.5: Conservation of genes from *C. concisus* 13826 plasmids across RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17 using orthoVenn.

<table>
<thead>
<tr>
<th><em>C. concisus</em> strains</th>
<th>Orthologs</th>
<th>No hits</th>
<th>Orthologs</th>
<th>No hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCCON31 total genes 33</td>
<td>pCCON16 total genes 23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUS22-Bd2</td>
<td>3</td>
<td>30</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>RMIT-JF1</td>
<td>3</td>
<td>30</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>RMIT-O17</td>
<td>3</td>
<td>30</td>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td>RCH 26</td>
<td>16</td>
<td>17</td>
<td>7</td>
<td>16</td>
</tr>
</tbody>
</table>
Six orthologous hits from pCCON16 included a hypothetical protein (EAT97364.1), mobilization protein (EAT97353.1) and four conserved hypothetical proteins (EAT97344.1, EAT97362.1, EAT97363.1 and EAT97346.2). In RMIT-O17 two orthologous hits with pCCON16 are EAT97364.1 and EAT97346.2, which were also present in AUS22-Bd2 and RMIT-JF1. Three hits with pCCON31 were the same as AUS22-Bd2 and RMIT-JF1. RCH 26 had 7 orthologous hits with pCCON16, which were same as AUS22-Bd2 and RMIT-JF1 and additionally Cpp12 (EAT97358.1). RCH 26 had 16 orthologous genes with pCCON31. Further analysis revealed that 4 of these genes were from contig 11 of the assembly which are all hypothetical proteins while the rest of the orthologous genes were from contig 21 and contig 17. Interestingly, contig 17 and contig 21 were found to be located in the same plasmid in RCH 26. This is discussed in detail in the next section.

5.3.6 Plasmid detection, mapping and construction

A feature of replication origins which is common to all methods of replication is the presence of directly repeated sequences to which replication proteins bind. In addition there may be an adjacent (A+T) rich region containing repeats and one or more dnaA boxes where the DnaA initiator protein binds (del Solar et al., 1998). No such repeated sequence was detected in AUS22-Bd2, RMIT-O17 and RMIT-JF1 when analysed with the web based tool ‘Tandem Repeat Finder’ (Benson, 1999). Only RCH 26 showed two contigs, (contig 17 and contig 20) to contain tandem repeats by using ‘Tandem Repeat Finder’. Contig 20 contained four repeats of a 22 bp long sequence ‘TATTCAAGGATAGATTTCTCAC’ followed by an (A+T) rich region. Contig 17 contained 3.5 repeats of an 11 bp long sequence ‘AATATGATAAT’ followed by an (A+T) rich region. Both of these contigs could amplify PCR products by inverse PCR,
using the primer sets described in Table 5.1. The inverse PCR products of contig 17 (~2400bp) and contig 20 (~2800bp) are shown in Figure 5.7. In the assembly performed by CLC genomic workbench the average coverage of contig 20 was 3,015.94x while contig 17 had average coverage only 344x. The average coverage of the assembly of RCH 26 was 555x (Table 5.4). The plasmid containing contig 20 was named as pCCON20 with a size of 3,316 bp and the plasmid containing contig 17 was named as pCCON22 with a size of 22,523 bp.

5.3.6.1 Characteristics of pCCON20

The inverse PCR product amplified from contig 20 was sequenced by Sanger sequencing method using the primers used for the PCR. After analysing the sequenced data, the full plasmid was constructed (Appendix V) and annotated. The Sanger sequencing data and the inverse PCR revealed that contig 20 is a plasmid itself with a size of 3,316 bp and %GC content of 37.4. Three CDS were predicted in the plasmid (Figure 5.8A) and named clockwise in the map as: i) pCCON20.1, homologous with the RepB family plasmid replication initiator protein from C. hominis with 97% query cover and 40% identity; ii) pCCON20.2: homologous with hypothetical protein of Betaproteobacteria bacterium with 86% query cover and 32% identity; and iii) pCCON20.3, identified as the Mobilization protein BmpH, with 78% query cover and 32% identity with mobilization protein of C. coli.
Figure 5.7: Inverse PCR performed on RCH 26 gDNA to confirm plasmids in contig 17 and contig 20 showing the size of PCR products by gel electrophoresis. Lanes 1: PCR product by primers from contig 17 (~2400bp); Lane 2: -ve control; Lane 3: PstI digested λ DNA ladder; Lane 4: PCR product by primers from contig 20 (~2800bp)
5.3.6.2 Characteristics of pCCON22

The inverse PCR product from contig 17 produced a larger size product than expected. After Sanger sequencing using the inverse PCR primers it revealed that another small contig (contig 21) was included in the whole plasmid with contig 17. When two contigs were combined following the Sanger sequence, a whole plasmid was constructed with 22,523 bp length and a lower %GC compared to the previous plasmid, pCCON20 (31.8%). A total of 26 CDS were predicted in the annotation and they were clockwise named from pCCON22.1–pCCON22.26 (Figure 5.8B). However, 19 of these CDS were hypothetical proteins and only seven CDS were positively identified. The identified proteins were: incQ plasmid conjugative transfer DNA nicking endonuclease TraR (pTi VirD2 homolog); DNA primase; Chromosome (plasmid) partitioning protein ParA; Channel-forming transporter/cytolysins activator of TpsB family; Channel-forming transporter/cytolysins activator of TpsB family; and 2 copies of a putative large exoprotein involved in heme utilization or adhesion of ShlA/HecA/FhaA family. Interestingly, eight proteins from this plasmid were homologous with proteins of a plasmid (pCCON31) of C. concisus 13826.

5.3.7 Pan- and core-genome analysis

The core and pan genomes of six C. concisus strains (AUS22-Bd2, RMIT-O17, RMIT-JF1, RCH 26, C. concisus 13826 and ATCC 33237T) were defined using BPGA 1.2 pipeline (Bacterial Pan Genome Analysis Tool) (Chaudhari et al., 2016). Using 50% sequence identity as a cut-off value BPGA 1.2 extracted 1463 genes as a core genome from a pool of 11,728 genes from six C. concisus genomes.
Figure 5.8: A circular representation of plasmid sequences present in RCH26 genome. A: pCCON20 (3,316 bp), B: pCCON22 (22,523 bp). The colours are explained in the box at the top.
Chapter 5: Comparative genomics of oral and intestinal *C. concisus*

 Genome wise statistics i.e., distribution of core, accessory, unique, and exclusively absent genes (absent in a particular strain but present in others) have also been generated by BPGA 2.1 pipeline and presented in Table 5.6. After adding the representative core, accessory, and unique gene sets, the pan genome was found to consist of 2,790 genes.

 The pan genome profile curves are shown in Figure 5.9. The empirical power law equations and exponential equations (Table 5.7) were used for extrapolation of the pan and core genome curves respectively, as displayed in the respective plots generated by the clustering tool. The pan-genome curve has almost reached a plateau, indicating that the *C. concisus* pan-genome would probably close in the near future.

 BPGA 2.1 pipeline was used for COG (Clusters of Orthologous Groups of proteins) function and KEGG (Kyoto Encyclopedia of Genes and Genome) pathway mapping of representative protein sequences of core, accessory and unique clusters of the *C. concisus* strains in this study. Figure 5.10 shows the COG distribution profiles of the core, accessory, and unique gene families. Detail COG distribution among these strains is presented in Figure 5.11. The COG distribution pattern showed involvement of more core genes in amino acid transport and metabolism; energy production and conversion and translation; ribosomal structure and biogenesis. Accessory and unique genes appear to be enriched in replication, recombination and repair, and cell wall/membrane/envelope biogenesis related functions. The accessory and unique genome had more hypothetical genes compared to the core genome (data not shown).
The KEGG distribution pattern (Figures 5.12 and 5.13) indicated involvement of more core genes in cell motility, energy metabolism, cofactors and vitamin metabolisms, nucleotide metabolism and translation, while accessory and unique genes appeared to be enriched in carbohydrate metabolism only. Unique genes are more involved in amino acid metabolism and signal transduction. The KEGG pattern also showed both core and unique genes were involved in cell growth and death, glycan biosynthesis and metabolism, metabolism of terpenoids and polyketides, nucleotide metabolism and replication and repair. Interestingly, there was no unique gene associated with human diseases including cancers, cardiovascular diseases, endocrine and metabolic diseases, infectious diseases, immune diseases and neurodegenerative diseases (Figure 5.13).

Two phylogenetic trees were generated by BPGA 1.2 pipeline. One was based on concatenated core gene alignments and the other one using pan-matrix (i.e., binary gene presence/absence (1/0) matrix). These two types of C. concisus strains trees generated using USEARCH clustering are shown in Figure 5.14 (A and B). As expected, there are some variations across the trees generated by different types of phylogenetic approaches. In the core phylogenetic tree C. concisus 13826 and ATCC 33237T were the most closely related while RMIT-O17 was the most distant strain. The phylogenetic tree generated from the pan genome also showed C. concisus 13826 and ATCC 33237T in the same clade; however, the most closely related strains were RMIT-JF1 and RCH 26. Again, in this tree RMIT-O17 remained the most distant strain.
Table 5.6: Distribution of core, accessory, unique and exclusively absent genes among
*C. concisus* 13826, ATCC 33237, RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17 using
USEARCH clustering tool.

<table>
<thead>
<tr>
<th>Genome no.</th>
<th><em>C. concisus</em> strains</th>
<th>No. of core genes</th>
<th>No. of accessory genes</th>
<th>No. of unique genes</th>
<th>No. of exclusively absent genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>C. concisus</em> 13826</td>
<td>1463</td>
<td>206</td>
<td>290</td>
<td>63</td>
</tr>
<tr>
<td>2</td>
<td>ATCC 33237&lt;sup&gt;T&lt;/sup&gt;</td>
<td>1463</td>
<td>237</td>
<td>31</td>
<td>24</td>
</tr>
<tr>
<td>3</td>
<td>AUS22-Bd2</td>
<td>1463</td>
<td>260</td>
<td>88</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>RCH 26</td>
<td>1463</td>
<td>320</td>
<td>171</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>RMIT-JF1</td>
<td>1463</td>
<td>402</td>
<td>104</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>RMIT-O17</td>
<td>1463</td>
<td>301</td>
<td>101</td>
<td>6</td>
</tr>
</tbody>
</table>
Chapter 5: Comparative genomics of oral and intestinal *C. concisus*

Table 5.7: Statistical analysis of the pan and core genome plot of *C. concisus* 13826, ATCC 33237, RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17.

<table>
<thead>
<tr>
<th></th>
<th>PAN GENOME</th>
<th>CORE GENOME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fit law</td>
<td>POWER</td>
<td>EXPONENTIAL</td>
</tr>
<tr>
<td>Equation</td>
<td>$f(x)=a.x^b$</td>
<td>$f(x)=c.e^{(d.x)}$</td>
</tr>
<tr>
<td>Parameters</td>
<td>$a= 1850.62$</td>
<td>$c= 1840.85$</td>
</tr>
<tr>
<td></td>
<td>$b= 0.220262$</td>
<td>$d= -0.0508756$</td>
</tr>
<tr>
<td>Expected Size</td>
<td>2790</td>
<td>1463</td>
</tr>
<tr>
<td>Estimated Size</td>
<td>2746.09</td>
<td>1356.59</td>
</tr>
</tbody>
</table>

The parameter 'b' = 0.220262, the pan genome is still open but may be closed.

Figure 5.9: Core and pan genome plot of six *C. concisus* strains using clustering tools-USEARCH.
Chapter 5: Comparative genomics of oral and intestinal *C. concisus*

Figure 5.10: COG distribution of core, accessory and unique genes in *C. concisus* 13826, ATCC 33237, RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17.
Figure 5.11: Detail COG distribution of core, accessory and unique genes in *C. concisus* 13826, ATCC 33237, RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17.
Figure 5.12: KEGG distribution of core, accessory and unique genes in *C. concisus* 13826, ATCC 33237, RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17.
Figure 5.13: Detail KEGG distribution of core, accessory and unique genes in *C. concisus* 13826, ATCC 33237, RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17.
Figure 5.14: Phylogenetic analysis by BPGA 1.2 pipeline using genomes of *C. concisus* 13826, ATCC 33237, RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17 based on: (A) concatenated core genes (B) binary pan-matrix. The distance of branches is shown by the scale below.
5.3.8 Enrichment analysis using blast2GO

Functional enrichment analysis of core and pan genomes of six selected *C. concisus* strains using Blast2GO is presented in Figure 5.15. The analysis was performed by core genome as test set and pan genome as reference set. Several gene ontologies were found to be enriched within two sets of genes. The core genome was enriched or over-represented in GO terms of ribosome biogenesis; rRNA binding; tRNA modification; many metabolic processes (purine, pyrimidine, pyruvate and amino acid); and many biosynthetic processes (amino acid, purine, pyrimidine, phospholipid, nicotinamide and peptidoglycan). Different ‘metal-ion binding’ proteins (zinc ion binding, magnesium ion binding, iron-sulphur cluster binding) were also enriched in the core genome.

In the pan genome, GO terms of defence response, clearance of foreign intracellular DNA and other DNA-related processes such as DNA restriction and modification, DNA integration and DNA recombination were enriched (over-represented) (Figure 5.15). Moreover, C-5 methylation of cytosine and DNA (cytosine-5)-methyltransferase activity, Type 1 site-specific deoxyribonuclease complex and activity were enriched in pan genome of selected *C. concisus* strains. Integral components of cell membrane were also over-represented compared to the core genome (Figure 5.15).
Chapter 5: Comparative genomics of oral and intestinal *C. concisus*

**Figure 5.15:** Enrichment of Gene Ontologies for core and pan genes across *C. concisus* 13826, ATCC 33237, RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17 genomes using Fisher’s Exact Test.
5.3.9 Detection of orthologous clusters in four *C. concisus* genome

OrthoVenn clustering was applied to identify gene clusters enriched in four *C. concisus* genomes of RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17. The analysis showed that the four genomes contained a total 1913 clusters and 1,906 orthologous clusters (present in at least two strains) based on the protein sequences. The clusters contained between eight to two genes. The numbers in the Venn diagram (Figure 5.16) represent the number of orthologous clusters that each strain shares with the three other strains. The diagram shows that 1,593 gene clusters are shared by all four strains, suggesting their conservation in the lineage. Among these 1,593 clusters, 1,573 were single-copy gene clusters. These clusters are most likely gene clusters within multiple genes or in-paralog clusters. The presence of the in-paralog clusters suggests that there might be a lineage specific gene expansion in these gene families in that strain. Based on GO annotation of these clusters, some of these lineage specific clusters could be potentially involved in important biological processes. The number of total proteins, clusters and the singletons (not in any cluster and only one copy in the genome) are presented in Table 5.8. The shared clusters were investigated. Most of the proteins were found to be hypothetical. RCH 26 and RMIT-O17 shared capsular polysaccharide biosynthesis protein CapD along with other hypothetical proteins which may be related to virulence. RCH 26 and RMIT-JF1 shared the virulence factor-related M protein. Interestingly, Type II secretion system protein D was shared by RCH 26, RMIT-JF1 and RMIT-O17, while RCH 26, AUS22-Bd2, RMIT-JF1 shared Methyl-accepting chemotaxis protein 4 which helps bacteria to swim towards nutrients and away from toxins. Interestingly, RCH 26 has the most singletons while AUS22-Bd2 has the least.
Figure 5.16: Venn diagram generated using OrthoVenn web based program showing the distribution of shared gene families (orthologous clusters) among RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17. The cluster number in each component is listed in the graph below.
Table 5.8: Counts of total proteins, clusters and singletons present in AUS22-Bd2, RCH 26, RMIT-JF1 and RMIT-O17 genomes.

<table>
<thead>
<tr>
<th>Species</th>
<th>Proteins</th>
<th>Clusters</th>
<th>Singletons</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUS22-Bd2</td>
<td>1859</td>
<td>1732</td>
<td>118</td>
</tr>
<tr>
<td>RCH 26</td>
<td>2141</td>
<td>1785</td>
<td>343</td>
</tr>
<tr>
<td>RMIT-JF1</td>
<td>2048</td>
<td>1845</td>
<td>188</td>
</tr>
<tr>
<td>RMIT-O17</td>
<td>1902</td>
<td>1759</td>
<td>135</td>
</tr>
</tbody>
</table>
5.3.10 Genes/proteins associated with virulence

An investigation into the presence of some virulence related genes or proteins, such as Zot, phospholipases A, S-layer RTX or NhaC and exotoxin 9, putative virulent genes such as cjaA, cjaC, dnaJ and motility related genes flaA, flab and flaC, was performed in the genomes of AUS22-Bd2, RCH 26, RMIT-JF1 and RMIT-O17. Table 5.9 represents the presence of virulence genes in four C. concisus strains. Interestingly, Zot which increases intestinal permeability by targeting the tight junction was found in RCH 26 and RMIT-JF1 only. Two virulence factors S-layer RTX, phospholipase A, the flagella genes and other putative virulence genes were found in all C. concisus strains. However, four invasion related proteins were absent in RMIT-O17, a strain that was isolated from a healthy person. This finding supported the phenotypic characteristics found previously (Section 3.6.3.3), where invasive index of RMIT-O17 was detected as 0.281 and suggested as nonenteric-invasive C. concisus (EICC) strain.

As the zot was only found in the genome sequences of RCH 26 and RMIT-JF1, more investigation was performed. The multiple sequence alignment (MSA) of zot in RCH 26 and RMIT-JF1 along with C. concisus 13826 zot is shown in Figure 5.17. Previously a polymorphism in the zot (substitution of valine at amino acid position 270 of C. concisus 13826) has been reported in C. concisus associated with active IBD patients (Mahendran et al., 2013). However, no substitution of valine at position 270 was observed in RCH 26 and RMIT-JF1, isolates from gastroenteritis and Crohn’s disease patients, respectively, as indicated in Figure 5.17 where ‘aa’ position 270 is indicated in a box.
Table 5.9: Presence of virulence related genes among RCH 26, AUS22-Bd2, RMIT-JF1 and RMIT-O17 genomes.

<table>
<thead>
<tr>
<th>C. concisus strains</th>
<th>AUS22-Bd2</th>
<th>RCH 26</th>
<th>RMIT-JF1</th>
<th>RMIT-O17</th>
</tr>
</thead>
<tbody>
<tr>
<td>zot</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>S-layer RTX/ T1SS secreted agglutinin RTX</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Outer membrane phospholipase A</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Adhesion related proteins</td>
<td>NhaC</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>NhaA</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Invasion related proteins</td>
<td>MobA</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Exotoxin 9</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>StbE replicon stabilization toxin</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RelB/StbD replicon stabilization protein (antitoxin to RelE/StbE)</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Putative virulent genes</td>
<td>cjaC</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>cjaA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>dnaJ</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Flagellar genes</td>
<td>flaA</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>flaB</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>flaC</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
Chapter 5: Comparative genomics of oral and intestinal *C. concisus*

CLUSTAL O(1.2.2) multiple sequence alignment

<table>
<thead>
<tr>
<th></th>
<th>zot_13826</th>
<th>RCH_26</th>
<th>RMIT-JF1</th>
</tr>
</thead>
<tbody>
<tr>
<td>216</td>
<td>MSLIIGPRSGKTYAVHLINDYELHLKGGESKRYFYTNYINGKLDFDHFDFGVQYDKN</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>216</td>
<td>DFLTAVSQEYTLSSQYENGFLVDVNYDEVALKGIVYHCLIVLDEAYNTFTKTFND</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>216</td>
<td>SLGRFLSYHGFIDILIFQSRQTRNREYLVHTELMMAQPSGRKLSLKFKYKVFYSTS</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td>216</td>
<td>SQVNDNINSENKLFFQKISNLYSSGSNEIYKSYATKinkiFFALAFVFSYVVKFLEPKH</td>
<td>240</td>
<td></td>
</tr>
<tr>
<td>216</td>
<td>EPAQSTQETRFVDLNASDKNIKAISSNREKSDINTTIFNDKIYLRTCPFSGCKFKN</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>216</td>
<td>YAIDLSLSFLELLSSSNCHIFLHDKKSNGYIDYFVGSCNFVRVLRKGNENSQRFVNEK</td>
<td>360</td>
<td></td>
</tr>
<tr>
<td>216</td>
<td>SPQTDSSMFPFTHK</td>
<td>374</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5.17: Multiple sequence alignment of *zot* found in RCH 26 and RMIT-JF1 and presence of valine at aa position 270 is indicated in the box.
5.4 Discussion

Four *C. concisus* strains were selected in this study to generate genomic read-data of which two were isolated from oral cavity and two from biopsy and stool samples, respectively. So far in the NCBI database the genome sequence of eight enteric *C. concisus* isolated from Crohn’s disease patients, gastroenteritis and faeces from healthy person, and oral isolate from gingival sulcus are available. In this project, two oral strains from healthy and Crohn’s disease patients were studied as no genome sequence available in the database from these categories. AUS22-Bd2 was selected for genome sequencing and further analysis, as it was the first *C. concisus* strain to be isolated from the duodenum, according to our knowledge, and may provide important new information.

In this project, two different next generation sequencers were used for sequencing four *C. concisus* strains. Sequencing of RCH 26 and RMIT-JF1 was performed by Ion proton next generation sequencer as at that time this was the only next generation sequencer at RMIT University, Melbourne. Ion Torrent and Illumina MiSeq can produce near perfect coverage behaviour on GC-rich, neutral and moderately AT-rich genomes, but Ion proton gives biased reading on extremely AT-rich genomes (Quail et al., 2012). As *C. concisus* is a moderate %GC bacterium (average %GC 39) and relatively smaller genome size, good quality data were expected from the Ion proton next generation sequencer. From Ion proton data report, high $\geq$ Q20 (base call accuracy of 99%) of these two genome sequences indicated good quality data (Table 5.3). However, the sequence industry is now dominated by Illumina which is preferable for bacterial genome sequencing considering the cost per Mb and errors rates (Bertelli and Greub, 2013). The Illumina Miseq was available at RMIT University in 2015 and was used to sequence the next two genomes of *C. concisus*, AUS22-Bd2 and RMIT-O17. Illumina Miseq produced very high quality data
as the $\geq$ Q40 (base call accuracy of 99.99%) of these two sequences were reasonably high (Table 5.4). However, the data produced by the two different next generation sequencers could not be compared as $\geq$ Q40 of Ion proton was not available. Different assembly methods such as Mira4, CLC genomic workbench and A5-miseq pipeline were used on availability. N50 (Size at which 50% of the genome is contained within contiguous sequences of this size or greater) in these four genome sequence assemblies were of high number indicating good quality of the assembly. The summed contig or scaffold lengths of these assemblies vary from 1.94 Mb to 1.83 Mb, with RMIT-JF1 having the largest genome and AUS22-Bd2 having the smallest. It is known that the genome size of \textit{C. concisus} can vary from 1.81 Mb to 2.11Mb across strains (Deshpande \textit{et al.}, 2013). The genome of RMIT-JF1 has highest number of genes and proteins and two copies of the \textit{rrn} (ribosomal RNA) operon which will be discussed in detail in Chapter 6.

Pairwise alignment using the Mauve alignment tool revealed very interesting findings. The four \textit{C. concisus} strains sequenced in this project showed very high level of gene shuffling and inversion when compared with the \textit{C. concisus} 13826 genome sequence. This was in agreement with the previous findings where pairwise comparison of seven \textit{C. concisus} strains with \textit{C. concisus} 13826 found to be much diversified but possessed higher similarity with ATCC 51561 (Deshpande \textit{et al.}, 2013). \textit{C. concisus} 13826 and ATCC 51561 are both genomospecies B strains, hence similar arrangements of genes were observed. On the other hand, pairwise comparison with ATCC 33237$^T$ showed a very high level of similarity with four of \textit{C. concisus} genome sequenced in this study. ATCC 33237$^T$ and four strains sequenced in this study are all from genomospecies A which explains the higher similarity by pairwise comparison. It has been suggested that \textit{C. concisus} 13826 may not be a good representative of the species (Kaakoush \textit{et al.}, 2011a)
and this is also supported by the analysis of pairwise comparison with four *C. concisus* genomes in this study (Figure 5.3). The comparative genomic circular view based on sequence based comparison of the four sequenced *C. concisus* strains with *C. concisus* 13826 and ATCC 33237T using Seed viewer tools revealed that there was more diversity when compared to *C. concisus* 13826 than ATCC 33237T. This finding was in line with the previous result of pairwise comparison. However, prominent gaps in gene content were observed when the four genomes were compared to ATCC 33237T.

Two bands for possible plasmids were observed in the agarose gel electrophoresis of RCH 26 along with the prominent chromosomal DNA band, while the other three *C. concisus* strains had only one chromosomal band. This finding was consistent when plasmids were searched for ‘ori’ region by the ‘Tandem Repeat Finder’ as only two contigs contained ‘ori’ region. In addition, the PCR product obtained using the inverse PCR confirmed the existence of the plasmid (Figure 5.7). The smaller plasmid in RCH 26, pCCON20 is unique, the presence of mobilization protein BmpH and RepB (a replication protein) also confirmed this contig to be a plasmid. The high coverage number of pCCON20 indicated that it is a high copy number plasmid. Furthermore, no significant similarity was found in the database when nucleotide BLAST was done. The annotation of the plasmid showed it contains three open reading frames. However, none of those were homologous with any *C. concisus* gene but were homologous with other enteric pathogens including *C. coli* and *C. hominis* and another Gram negative enteric bacteria *Betaproteobacteria bacterium*. This finding indicates that RCH 26 had acquired pCCON20 from the intestine via horizontal transfer from other intestinal inhabitants. Because of the small size and high copy number, this plasmid was selected for constructing the complementary luxS plasmid discussed in Chapter 4. The other plasmid,
pCCON22 was larger and appeared in two contigs from the assembly. This plasmid had 51% query cover and eight genes including a hemolysin gene and Outer membrane toxin activator protein CdiB-2 shared with C. concisus 13826 plasmid pCCON31. However, the shared genes present in pCCON22 are not contiguous i.e. gene shuffling was observed. The presence of Hemolysin and CdiB-2 in the plasmid is an indication of the virulence properties of C. concisus RCH 26. CdiB is an outer membrane pore-forming protein allowing secretion of the CdiA polymorphic toxin, a large filamentous protein (Aoki et al., 2010) involved in heme utilization or adhesion of ShlA/HecA/FhaA family. Secreted haemolytic activity in C. concisus was reported previously by Istivan et al. (2008). In summary, it can be concluded that plasmids found in C. concisus vary considerably in their identity, conservation, replication methods and function.

Pan and core genome analysis was performed using 50% cut-off value and 1,463 core and 2,790 pan genes were obtained from six C. concisus strains in this study. Two reference C. concisus genome sequences were included in this analysis to increase the validity of the test. However, in a previous study by Deshpande et al. (2013), a larger core and pan genome (1,556 and 3,254 genes respectively) was defined with 40% homology and 40% length from eight C. concisus genomes indicating that the number of defining core and pan genome can vary depending on the bioinformatics pipeline and cut-off value. The advantage of using BPGA 1.2 pipeline was availability of the accessory and unique gene list for strains included in the analysis. Involvement of more core genes in amino acid transport and metabolism, energy production and conversion and translation, ribosomal structure and biogenesis were an indication of the validity of the analysis. More hypothetical genes and genes related to repair and replication in unique and accessory genomes suggested that the virulence genes were included in the pan genome i.e. as
accessory or unique genes. This was also supported by the enrichment analysis by Blast2GO, where the enrichment of RNA processes, metabolic and biosynthetic processes were observed within the core genome of *C. concisus* which was expected. In the core genome the ‘metal-ion binding’ GO terms were enriched as many metabolic enzymes have metal binding catalytic centres. On the other hand, enrichment of defence responses, DNA-related processes such as DNA integration and DNA restriction-modification in the pan genome suggests that different defence mechanisms exist in the efficiency of these *C. concisus* strains to probably survive phage attacks (Seed, 2015).

In this study, the shared or orthologous genes of *C. concisus* were investigated. The genes shared by six *C. concisus* strains including the four *C. concisus* sequenced at RMIT University and the two reference *C. concisus* genomes were analysed by detecting pan and core genome. OrthoVenn was used for detecting genes which were shared between two or three strains. However, most of the genes shared between any of these strains were hypothetical genes therefore no further information could be obtained. From this analysis, it was revealed that RCH 26, AUS22-Bd2, RMIT-JF1 are sharing the most number of genes including chemotaxis protein. AUS22-Bd2, RMIT-JF1 and RMIT-O17 shared only 26 clusters but containing antiporter NhaA which has sodium ion transmembrane transporter activities. Previously, this protein was identified and correlated with *C. concisus* low adherence (Deshpande *et al.*, 2013). Presence of prophage CP4-57 integrase in RCH 26 and RMIT-JF1 indicated a viral entry into the host cell. The virulence-related M protein shared by these two strains indicated that they have the mechanism to escape the phagocytosis process. The M protein is an important virulence factor of group A streptococci (GAS) and is essential for providing antiphagocytic functions critical to survival in human tissues and fluids (Metzgar and Zampolli, 2011).
Two strains isolated from IBD patients in our study (RMIT-JF1 and AUS22-Bd2) shared the FeMo cofactor biosynthesis protein NifB, which has a catalytic activity in nitrogen fixation. These two strains also contain uncharacterized membrane protein YabM, which is an integral membrane component engaged in the polysaccharide biosynthetic process and peptidoglycan synthesis and may play a role in virulence.

The annotated genomes of the four C. concisus strains were screened for virulence associated genes or proteins. The RMIT-O17 strain was found it to be adhesive but non-EICC (Section 4.5.3.3). From the genome analysis it was found that this strain has the adhesion related proteins (NhaC and NhaA), but not the four invasive related proteins (MobA, Exotoxin 9, StbE replicon stabilization toxin and RelB/StbD replicon stabilization protein). Furthermore, NhaC was associated with high adherence and NhaA was associated with low adherence of C. concisus strains (Deshpande et al., 2013). However, in this study, both were found in the genome of all strains except the RCH 26 genome, where NhaA was absent. The absence of Zot in RMIT-O17 correlates with its isolation source from a healthy individual. Zot was found in RCH 26 and RMIT-JF1, from a gastroenteritis patient and Crohn’s disease patient respectively. Polymorphism (valine substituted by leucine) in Zot was investigated in this protein, however, the presence of valine in both of the proteins at aa position 270 contradicts with the previous finding (Mahendran et al., 2013), where a substitution of valine at position 270 was significantly associated with patients with active IBD, but not in healthy controls ($p = 0.011$). The AUS22-Bd2 strain has been screened, and the strain was found to be invasive, exhibiting an invasion index of 3.5 (data not shown). The presence of exotoxin 9 and other invasion related genes in AUS22-Bd2 also correlates with the invasiveness of this strain and is also in agreement with the source of isolation of this strain (duodenum of an IBD patient).
Chapter 5: Comparative genomics of oral and intestinal *C. concisus*

RCH 26 has been assayed for adhesion and invasion by Elshagmani (2015), and the strain was found to be non-invasive (index invasive value <1). The RMIT-JF1 has not yet been assessed for invasion and adherence. However, the all but one invasion related genes/toxins to were found to be present in these strains (StbE replicon stabilization toxin was absent in both RMIT-JF1 and RCH 26), in keeping with the known virulence associated with these strains (isolated from oral cavity of a Crohn’s disease patient and the stool of a patient with gastroenteritis, respectively).

Taken together, a genomic comparison has been performed between *C. concisus* strains isolated from oral and intestinal sources with different clinical histories. From this analysis it can be concluded that there is a significant difference in gene content depending on the isolation site and also on the clinical history of the host. The presence of plasmids in the genome can vary in *C. concisus* and none were found in oral isolates. The genes present in the analysed plasmids were also highly variable. Moreover, the presence of a virulence related gene in the genome does not appear to confer virulence, at least using the standard assays in this study. Further analysis and the expression of the virulence genes are suggested as future direction for this study. Phylogenetic analysis is an important approach for investigation of evolutionary relationships, which is discussed in detail in the next Chapter.
Chapter Six

Analysis of the ribosomal RNA (rrn) operon of *Campylobacter concisus*

6.1 Introduction

*C. concisus* is a heterogeneous species, consisting of two or more genomospecies which are phenotypically indistinguishable but genetically distinct taxa (Vandamme *et al.*, 1989; Bastyns *et al.*, 1995; On and Harrington, 2000; Istivan *et al.*, 2004; Aabenhus *et al.*, 2002a; Aabenhus *et al.*, 2005a; Engberg *et al.*, 2005). The 23S rRNA gene has been utilized widely for rapid detection and identification of *C. concisus* (Bastyns *et al.*, 1995; Istivan *et al.*, 2004; Kalischuk and Inglis, 2011). *C. concisus* strains can be divided into two molecular groups (genomospecies A and B) by PCR amplification of the 23S rDNA region (Istivan *et al.*, 2004; Kalischuk and Inglis, 2011). Although, the 16S rRNA gene has been a useful marker to study phylogenetic relationships and for detection/identification in bacteriology (Woo *et al.*, 2008), there is no report on utilizing the 16S rRNA gene as a target to classify *C. concisus* so far. However, it has been used to detect other *Campylobacter* spp. (Linton *et al.*, 1997; Maher *et al.*, 2003). The internal transcribed spacer (ITS) region that lies between the 16S and 23S rRNA subunit genes has been utilised for many Gram negative bacteria for identification and differentiation (Gürtler and Stanisich, 1996). The variation in sequence in the ITS region has been used to detect and distinguish some thermotolerant *Campylobacter* spp. including *C. jejuni*, *C. coli* and *C. lari* (Khan and Edge, 2007; Hayashi *et al.*, 2012). There was evidence of the presence of strain-specific intervening sequences (IVS) within the 23S rRNA gene of *C. jejuni* and *C. coli* (Man *et al.*, 2010b). Currently, there are very few reports on the ITS region in *C. concisus* (Khan and Edge, 2007; Man *et al.*, 2010b). The diversity of the *rrn*
Chapter 6: Analysis of *rrn* operon in *C. concisus*

(ribosomal RNA) operon region observed previously in *Campylobacter* spp. (Khan and Edge, 2007; Hayashi *et al.*, 2012; Man *et al.*, 2010b) suggests that this region may be a useful tool for typing *C. concisus*.

The aim of this study was therefore to

1. Investigate the 5S rRNA gene, 16S rRNA gene, 23S rRNA gene and ITS regions (*rrn* operon) of selected *C. concisus* isolates.

2. Differentiate and establish a systematic relationship among the selected *C. concisus* strains.

### 6.2 Materials and methods

#### 6.2.1 Source of *rrn* sequences

The *rrn* sequences were analysed from the following strains (Table 6. 1):

**Collection (A):** consisted of *rrn* operons from four *C. concisus* strains described in Chapter 5. These were: RCH 26, RMIT-JF1, RMIT-O17 and AUS22-Bd2. The *rrn* operons were extracted from the whole genome sequence assembly. Two alleles of *rrn* operons were detected in RMIT-JF1 and both were included in this analysis.

**Collection (B):** consisted of the reference genomes *C. concisus* 13826 and *C. concisus* ATCC 33237T and sequences extracted from whole genome shotgun sequence data from the NCBI database. Six of the isolates (*C. concisus* ATCC 51561, *C. concisus* ATCC 51562, UNSWCS, UNSW1 and UNSW3) (Deshpande *et al.*, 2013) and UNSWCD (Deshpande *et al.*, 2011) have been sequenced by Illumina HiSeq sequencing platform. The UNSW2 genome (Deshpande *et al.*, 2013) was excluded from this study as the *rrn*
operon is divided into two contigs and sequences of the ITS region were missing. *C. concisus* 13826 has three alleles of *rrn* operons; all were included in the analysis as individual operon for analysis. *C. concisus* ATCC 33237T has three completely identical alleles of *rrn* operon; hence one was included in the analysis.

### 6.2.2 In silico analysis of *rrn* operons in *C. concisus*

Geneious Pro (version 8.1.7, www.geneious.com) bioinformatics software (Biomatters) was used to construct *rrn* operon sequence alignments, contigs and consensus sequences. Pairwise distance was measured by MEGA7 (Tamura *et al*., 2013). Statistical analysis of DNA sequence metadata was performed by the Ribotyping App (Gürtler *et al*., 2014). Fisher’s exact test was performed using GraphPad Prism 6 statistical analytical tool.

### 6.2.3 Phylogenetic analysis based on ribosomal RNA genes

Phylogenetic analysis was performed on the whole *rrn* operon and 16S and 23S rRNA genes separately. The evolutionary history was inferred using the Neighbor-Joining (NJ) method (Saitou and Nei, 1987). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) were shown next to the branches (Felsenstein, 1985). The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura *et al*., 2004) and are in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA7 (Tamura *et al*., 2013).
6.3 Results

6.3.1 Intraspecies sequence variability of rRNA genes

The 5S rRNA gene, 16S rRNA gene, 23S rRNA gene and ITS regions of 15 sequences from 12 strains of *C. concisus* were generated from online resources ([http://www.ncbi.nlm.nih.gov/genome/1439](http://www.ncbi.nlm.nih.gov/genome/1439)), namely *C. concisus* 13826 (3 alleles, *rrnA*, nucleotide position from 708209 to 714829, 6,192 bp in size; *rrnB*, nucleotide position from 1081796 to 1088098, 5,972 bp in size; and *rrnC*, nucleotide position from 1832146 to 1838220, 6,075 bp in size), *C. concisus* ATCC 33723 (5,742 bp), UNSWCD (5,516 bp), UNSWCS (5,584 bp), UNSW1 (6,002 bp), UNSW3 (5,699 bp), *C. concisus* ATCC 51561 (5,312 bp), *C. concisus* ATCC 51562 (5,659 bp) and four whole genome shotgun sequences performed at RMIT University: RCH 26 (5,651 bp), RMIT-JF1 *rrnI* (5,652 bp), RMIT-JF1 *rrnII* (5,106), RMIT-O17 (5,601 bp) and AUS22-Bd2 (6331bp), including three alleles of *C. concisus* 13826 and two alleles of RMIT-JF1(I and II) (Table 6.1). Each sequence was used to determine the variability among them. Average pairwise distance between sequences was 0.01. The shortest pairwise distance was between two alleles of *C. concisus* 13826, *rrnB* and *rrnC* which was 0.00019 and longest pairwise distance was 0.01839 between the sequences of ATCC 33237T and UNSWCS.

6.3.2 Indels in *rrn* operon of *C. concisus*

A total of 38 types of indels were identified when 15 sequences *rrn* operons were aligned in the Geneious Pro bioinformatics software (Figure 6.1 and Supplementary File). In the *rrn* operon, there are post-5S, 23S, 16S, ITS, and pre-16S regions. No IVS were found in the *rrn* operon in 15 sequences. In Figure 6.2, 38 indels including subtypes were plotted in a graph on the X-axis and each indel present in sequences were summed in the Y-axis.
Table 6.1: *C. concisus* strains used in this study, their genomospecies, *rrn* names, types and accession numbers.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genospecies</th>
<th><em>rrn</em> name</th>
<th><em>rrn</em> Type</th>
<th>Accession no</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMIT-JF1</td>
<td>A</td>
<td>A</td>
<td>I</td>
<td>JXUP0000000000</td>
<td>This study</td>
</tr>
<tr>
<td>RMIT-JF1</td>
<td>A</td>
<td>B</td>
<td>II</td>
<td>JXUP0000000000</td>
<td>This study</td>
</tr>
<tr>
<td>ATCC 51561</td>
<td>B</td>
<td>C</td>
<td>I</td>
<td>ANNH0000000000</td>
<td>(Deshpande <em>et al.</em>, 2013)</td>
</tr>
<tr>
<td>ATCC 51562</td>
<td>A</td>
<td>D</td>
<td>I</td>
<td>ANNI0000000000</td>
<td>(Deshpande <em>et al.</em>, 2013)</td>
</tr>
<tr>
<td>RCH 26</td>
<td>A</td>
<td>E</td>
<td>I</td>
<td>LVWL0000000000</td>
<td>This study</td>
</tr>
<tr>
<td>UNSW1</td>
<td>B</td>
<td>F</td>
<td>I</td>
<td>ANNF0000000000</td>
<td>(Deshpande <em>et al.</em>, 2013)</td>
</tr>
<tr>
<td>UNSW3</td>
<td>B</td>
<td>H</td>
<td>I</td>
<td>ANNE0000000000</td>
<td>(Deshpande <em>et al.</em>, 2013)</td>
</tr>
<tr>
<td>UNSWCD</td>
<td>B</td>
<td>I</td>
<td>I</td>
<td>AENQ0000000000</td>
<td>(Deshpande <em>et al.</em>, 2011)</td>
</tr>
<tr>
<td>UNSWCS</td>
<td>B</td>
<td>J</td>
<td>I</td>
<td>ANNG0000000000</td>
<td>(Deshpande <em>et al.</em>, 2013)</td>
</tr>
<tr>
<td><em>C. concisus</em> 13826</td>
<td>B</td>
<td>K</td>
<td><em>rrnA</em></td>
<td>CP000792</td>
<td>NCBI</td>
</tr>
<tr>
<td><em>C. concisus</em> 13826</td>
<td>B</td>
<td>N</td>
<td><em>rrnB</em></td>
<td>CP000792</td>
<td>NCBI</td>
</tr>
<tr>
<td><em>C. concisus</em> 13826</td>
<td>B</td>
<td>O</td>
<td><em>rrnC</em></td>
<td>CP000792</td>
<td>NCBI</td>
</tr>
<tr>
<td>RMIT-O17</td>
<td>A</td>
<td>P</td>
<td>I</td>
<td>LVLC0000000000</td>
<td>This study</td>
</tr>
<tr>
<td>AUS22-Bd2</td>
<td>A</td>
<td>Q</td>
<td>I</td>
<td>LVXF0000000000</td>
<td>This study</td>
</tr>
<tr>
<td>ATCC 33237(^T)</td>
<td>A</td>
<td>R</td>
<td><em>rrnA,B,C</em></td>
<td>CP012541</td>
<td>NCBI</td>
</tr>
</tbody>
</table>
6.3.3 Indels present in the post-5S region

Indel position 1, which is named as post-5S, was typed into two groups: i) post-5SA (light blue) and ii) post-5SB (light green). In indel type post-5SA there were 7 sequences from 5 strains (C. concisus 13826, ATCC 51561, UNSW1, UNSW3 and UNSWCS), and all are from genomospecies B. Indel-type post-5SB had 7 sequences from 5 strains sequenced in this study, C. concisus ATCC 51562 and C. concisus ATCC 33237T. In this sub-type all strains are genomospecies A. All of these sequences can be sub-typed.

Indel position 2 is in the ITS between 5S rRNA and 23S rRNA gene. All the sequences were typed into two groups: 5S-23SA and 5S-23SB. The former group was identified in all genomospecies A and the latter group was identified in all genomospecies B (\( p = 0.0003 \), by Fisher's exact test). Each type can be sub-typed into 6 different sub-types. This ITS is 145 bp in length for all but the second allele of RMIT-JF1 (rrnII), which has a single nucleotide deletion at position 683.

6.3.4 Indels present in 23S rRNA genes

A total of 21 indels were detected in 23S rRNA gene in 15 sequences. Indel position 3 has (TCTTGGCAGA) present in all genomospecies A strains and (CCTTGGCAGG) present in all genomospecies B strains (\( p = 0.0002 \), by Fisher's exact test). Indel position 4 was a single nucleotide change from G→A in both alleles of RMIT-JF1, ATCC 33237T, AUS22-Bd2 and RCH 26. Indel positions 4 to 11 and 14-20 had a single nucleotide polymorphism among these 15 sequences. Indel position 12 is a 55 bp long sequence in all genomospecies A and a different 55 bp long sequence in all genomospecies B (\( p = 0.0003 \), by Fisher's exact test). Indel position 13 is a 122 nucleotide
long sequence which can be typed as I23Ska and I23Skb that were detected in
genomospecies A and genomospecies B strains respectively ($p < 0.02$, by Fisher's exact
test). Interestingly, on the basis of the sequence this indel in genomospecies A can be
divided into three sub-types; I23Ska-1, I23Ska-2 and I23Ska-3. A total of 27 identical
SNP’s were present in all 7 strains in genomospecies A. The sequences of both alleles of
RMIT-JF1 and ATCC 33237T were identical within this 122 bp long indel. RMIT-O17 has
an SNP in position 1 which is the same as genomospecies B sequences. Additionally,
ATCC 51562 and RCH 26 have a SNP at position no 39 which places them in the second
sub-type I23Ska-2 as for the strain AUS22-Bd2. The third sub-type I23Ska-3 has a
different SNP at position no 75, which again is the same as genomospecies B sequences.
Indel position 21 is a 79 bp deletion in one allele of RMIT-JF1. Indel position 22 is
present in two types (Indel 23Sw and Indel 23Sx) present in genomospecies A and B
respectively. Indel 23Sw can be sub-typed in 4 groups. Indel 23 of 32 bp is present in all 8
sequences of genomospecies B and 2 alleles of RMIT-JF1. No IVS was found in the 23S
rRNA gene in 15 sequences analysed in this study.

### 6.3.5 Indels present in ITS region in rRNA genes

The ITS regions of all 15 sequences of *C. concisus* were highly variable in length
from 641 to 750 bp. The average length is 675 bp. The longest ITS is 750 bp in UNSWCS
and in UNSW1. The shortest ITS is 641bp in two alleles (*rrnB* and *C*) in *C. concisus*
13826. The average %GC content was 34.81 while the highest %GC content was 35.64 in
one allele in *C. concisus* 13826 (*rrnA*), while the lowest %GC content was 34.21 in ATCC
51562.
Chapter 6: Analysis of *rrn* operon in *C. concisus*
Figure 6.1: Map of ribosomal RNA (rrn) operons from 12 Campylobacter concisus strains showing pre-16S, 16S rRNA (red), ITS, tRNAile (light blue), tRNAala (orange), 23S rRNA (blue), 5S rRNA (purple) genes and post-5S regions (shown on top below the consensus). Below this is shown the percentage identity of the 15 sequences with green (100%), orange (75-100 %) and red (< 50 %). The grey bars represent the DNA sequences, black lines in the middle represent sequence gaps and coloured lines represent single nucleotide polymorphisms (SNPs) yellow, green, red and blue. The letters listed on the left hand side of each sequence represent the “rrn Type” in each strain with some strains containing more than one sequence. The coloured bars beneath each grey bar represent indels containing SNPs numbered 1-38.
Figure 6.2: Indel subtypes with genomospecies A and B from 15 sequences shown in Figure 6.1. Each position (1-38) shown on Figure 6.1 are marked at the bottom of each Indel subtype and add up to the number of sequences. The number of SNPs at each position varies according to whether the strain is genomospecies A or B.
A total of 3 indels were detected in the ITS sequences in 15 sequences at indel position no 24, 25 and 26. Indel no 24 was 130 bp long in all sequences and can be divided into two types, ITS1 indel a and ITS1 indel b. This indel is significantly associated with genomospecies A or B \((p = 0.0014, \text{ by Fisher's exact test})\). Indel type a contained all sequences from genomospecies B except UNSW3, which was in the second type with all 7 sequences from genomospecies A. ITS indel b can be sub-typed into 5 more divisions.

The second indel in the ITS sequences was at position 25 and it could be typed into two groups (ITS1 indel c and ITS1 indel d) and sub-typed into 4 and 6 sub-types respectively. In ITS1 indel c (173-174 bp long) there were four sequences, which were UNSW1, UNSW3, UNSWCS and one allele \((rrnA)\) from \(C. concisus\) 13826. ITS1 indel d consists of 65-67 bp long sequences in the rest of the sequences which were all from genomospecies A and the other two alleles \((rrnB\) and \(C)\) of \(C. concisus\) 13826, ATCC 51561 and UNSWCD.

The third indel in the ITS was at position 26 of a 12 bp long sequence \((TCTTATAGATAT)\) which was present in all sequences except for ATCC 51561 and the second allele of RMIT-JF1, while the indel \(TCTTATAGATAC\) was present in ATCC 51561. This indel was 11 bp long in the second allele of RMIT-JF1 with a deletion of T at the beginning of the indel and a single nucleotide change from \(T\rightarrow C\) at the end of the indel.

In all the sequences analysed, there were two tRNA genes found between indel 25 and 26; which were (I) tRNA \(^{\text{Ala}}\) (TGC) and (II) tRNA \(^{\text{Ile}}\) (GAT). The tRNA \(^{\text{Ala}}\) (TGC) was found at position 4583-4506 and tRNA \(^{\text{Ile}}\) (GAT) was found at position 4671-4595.
6.3.6 Indels present in 16S rRNA genes

A total of 9 indels were detected in the 16S rRNA gene in 15 sequences. In indel position 27 only RCH 26 had three SNPs and all other sequences were identical. At indel positions 28, 29, 31, 32 and 33 there were point deletions and substitutions. At indel 30 there was an insertion of (CATCTCTG) in UNSWCS instead of (TATCTCTA). At position 31 there was a SNP of T→C in all genomspecies A strains except for ATCC 51562. Indel 34 has an insertion of (TTAGCCTTGGTT) in the second allele of RMIT-JF1. Indel 35 is a 59 base long sequence (CTAGAGGAGGTAGAGTATTAGCAGTCGTTTCCACTGTGTCTCCTCTAGTGTTAGGCG) where a total of 5 subtypes were found in 15 sequences, whereas all the sequences from genomspecies A are in subtype I16Sr.

In the pre-16S rRNA there were 3 indels found. There was a 24 bp (CTTCTTTAGCGTAAAGCAAAGCTT) deletion in indel 36 in all sequences from genomspecies A and two alleles rrnB and rrnC of C. concisus 13826 from genomspecies B. Indel 37 was typed into two groups (Pre-16S2ai and Pre-16S2bi) and syb-typed into 5 groups. All genomspecies A belong to Pre-16S2bi group. Indel 38 is typed into two groups, Pre-16S3ai and Pre-16S3bi while Pre-16S3ai is found in all genomspecies A strains except for ATCC 33237T which has Pre-16S3bi as other genomspecies B strains.

6.3.7 Phylogenetic analysis of rrn operon of C. concisus

Sequences of the 15 rrn operons, 16S and 23S rRNA regions of 12 C. concisus strains were aligned to generate neighbor-joining (NJ) phylogenetic trees (Figures 6.3, 6.4 and 6.5). In each trees, two major clades were predicted. The distance scale is shown in
the figure. In one clade, the following sequences were present: *C. concisus* ATCC 33237\textsuperscript{T}, ATCC 51562, RCH 26, AUS22-Bd2, RMIT-O17 and two alleles of RMIT-JF1. Interestingly, all sequences from this clade are genomospecies A. In the other clade, all *C. concisus* sequences downloaded from the NCBI database, namely 3 alleles of *C. concisus* 13826, UNSWCD, UNSWCS, *C. concisus* ATCC 51561, UNSW3 and UNSW1 were found. In this clade all the sequences are from genomospecies B.
Figure 6.3: Evolutionary relationships of 15 sequences of \textit{rrn} operon from 12 \textit{C. concisus} strains. The optimal tree with the sum of branch length = 0.03504071 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The analysis involved 15 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 5138 positions in the final dataset.
Figure 6.4: Evolutionary relationship among 15 sequences of 16S rRNA gene from 12 C. concisus strains. The optimal tree with the sum of branch length = 0.01248230 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The analysis involved 15 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 1435 positions in the final dataset.
Figure 6.5:  Evolutionary relationship among 15 sequences of 23S rRNA gene from 12 *C. concisus* strains. The optimal tree with the sum of branch length = 0.03306552 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The analysis involved 15 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 2788 positions in the final dataset.
6.4 Discussion

In this study, the characteristics of the 5S rRNA gene, 16S rRNA gene, 23S rRNA gene, and the ITS region within 15 sequences (~6000 bp) of 12 *C. concisus* strains were investigated and their potential to be used in species differentiation and delineation of systematic relationships were evaluated. This is the first study to investigate the *rrn* operon of *C. concisus* by *in silico* analysis.

In general, the *Campylobacter* genome contains three copies of the *rrn* operon (Fouts *et al.*, 2005; Taylor *et al.*, 1992). This is unlike *E. coli*, with a genome almost three times the size of the *Campylobacter* genome that contains seven copies of the *rrn* operon (Krawiec and Riley, 1990). *C. jejuni* has three copies of the *rrn* operon (Kim and Chan, 1989) and all three operons possess a contiguous 16S-23S structure (Kim *et al.*, 1993). The same type of 16S-23S arrangement was found in the *C. concisus* genome and the complete whole genome sequences of *C. concisus* 13826 and ATCC 33237^T^ have been shown to contain three copies of the *rrn* operon. Interestingly, all three alleles were found to be completely identical in ATCC 33237^T^, while the three alleles of the *rrn* operon in *C. concisus* 13826 are contiguous but not identical. In the assembly of RMIT-JF1 WGS, two alleles of the *rrn* operon have been identified. In other 11 *C. concisus* shotgun sequences only one copy of the operon was identified (Table 6.1). The possible explanation is if the three alleles are identical, unlike *C. concisus* 13826, they could not be differentiated by shotgun sequencing. For the analysis it was presumed that all the shotgun WGS except RMIT-JF1 contain three identical copies of the *rrn* operon.

In the analysis 38 indels in the *rrn* genomic loci in 12 isolates of *C. concisus* (Figure 6.1) were identified. The 23S rDNA is much diversified yet all the indels are
systematically arranged in *C. concisus*. A total of 21 indels out of 38 indels identified in this study is from the 23S rRNA gene. In this analysis, it was found that subtypes of indels at positions 1, 2, 3, 12, 13 and 22 are present either in genomospecies A or genomospecies B. Interestingly, most of these indels are in 23S rRNA gene, only indels 1 and 2 are in the post-5S region. Indel 13 is a 122 bp long sequence and part of it was identified previously by Bastyns *et al.* (1995) when they designed a primer set (one forward –MUC 1-and two reverse primers –CON1 & CON2- in a mixture) from the 23S rDNA region for detection of *C. concisus* (Istivan *et al.*, 2004). Further, the authors grouped *C. concisus* into two genomospecies (genomospecies A and genomospecies B) using the reverse primers separately. MUC1 and CON1 amplified a product of 308 bp for genomospecies A while MUC1 and CON2 produced a product with similar size from genomospecies B. The analysis conducted in this study showed that the forward primer was designed for a conserved region of 23S rRNA gene and the two reverse primers were designed from indel 13. It can be concluded the PCR-based method for identification and classification of genetically highly diversified *C. concisus* used by Istivan *et al.* (2004) was convenient and rational.

A total of 9 indels were detected in the 16S rRNA gene in 15 sequences. Unlike 23S rRNA gene, the indels and the SNP are not specific to any genomospecies and this part of the *rrn* operon is more conserved compared to 23S rRNA gene (Figure 6.1). 16S rRNA has been used to identify and differentiate isolates belonging to the *Campylobacter*, *Arcobacter*, and *Helicobacter* genera previously by the PCR-RFLP method (Marshall *et al.*, 1999). Later Kulkarni *et al.* (2002) identified enteropathogenic campylobacters to the species level using the 16S rRNA gene by a PCR identification algorithm and PCR enzyme linked immunosorbent assay (PCR-ELISA). The 16S rRNA gene has been also
used for the detection and identification of Epsilobacteria by the PCR-DGGE method (Petersen et al., 2007). The authors identified 17 groups in Epsilobacteria by this method. Here three C. concisus isolates were assigned into two different DGGE groups without any consistency with genomospecies or source of isolation. Therefore, it can be concluded that the 16S rRNA gene may be very useful for identification, but not ideal for typing of C. concisus which is also supported by this analysis.

In this analysis it was shown that the ITS region is highly variable in C. concisus with 3 indels and two tRNAs. Christensen et al. (1999) investigated the internal spacer region (ISR) which is a synonym for ITS in C. jejuni and C. coli and reported the insertion of between 152 and 193 nucleotides in C. coli but not in C. jejuni in the ITS region. Later, the 23S rRNA gene and the ITS region within 30 taxa of the genus Campylobacter were investigated to use for species differentiation and delineation of systematic relationships (Man et al., 2010b). The authors reported that the ITS region produced the highest mean pairwise percentage difference (35.94%) compared to the 16S (5.34%) and 23S (7.29%) rRNA genes. C. jejuni and C. coli strains clustered in subgroups when the phylogenetic trees derived from ITS region, but this was not observed in Neighbor-joining trees derived from the 16S or 23S rRNA gene. It was suggested that the ITS region is the most appropriate region for the differentiation in rrn operon in Campylobacter spp.

The average %GC content of C. concisus was found as 34.81 in the ITS region in this study, which is in agreement with the finding by Man et al. (2010b), where they have found higher %GC contents (32.8 to 34.5%) in C. concisus compared to other Campylobacter spp (average 29.3%). However, Khan and Edge (2007) has found that the %GC content of C. concisus is considerably lower than other thermophilic Campylobacter
species such as *C. jejuni*, *C. coli* and *C. lari* and used this concept in a newly designed multiplex PCR from the ITS region of *Campylobacter* spp. where *C. concisus* was used as a negative control.

In this study no IVS were identified in the 15 sequences used for the analysis, neither in the 23S nor 16S rRNA genes. The presence of IVS is common in *Campylobacter* spp. as it has been found in the 23S rRNA gene in helix 49 in *C. jejuni*, *C. coli*, *C. curvus*, *C. fetus*, *C. upsaliensis* and *C. sputorum* biovar sputorum and fecalis (Konkel *et al.*, 1994; Meinersmann *et al.*, 2009; Man *et al.*, 2010b; Tazumi *et al.*, 2009). The presence of an IVS causes the fragmentation of 23S rRNA and the IVS can vary in size. Konkel *et al.* (1994) found a 157 bp long IVS in *C. jejuni* while Trust *et al.* (1994) found either a 147 or 37 bp IVS in *C. jejuni* and *C. coli*. IVS can be found in the 16S rRNA gene, since Linton *et al.* (1994) first reported its detection in *C. helveticus* in 1994. In another study, it was found that majority (66/69) of *C. coli* isolated from turkeys harboring IVSs in all three alleles of 23S rRNA genes were found to be resistant to erythromycin, and strains that had at least one IVS-free 23S rRNA gene were sensitive to erythromycin. Therefore the authors suggested that IVS can be acquired by these strains via natural transformation to erythromycin resistance (Chan *et al.*, 2007). Several researchers have investigated IVS in *C. concisus* and reported the absence of IVS in the *C. concisus* genome (Man *et al.*, 2010b; Tazumi *et al.*, 2009).

Furthermore, all 15 sequences were aligned and a phylogenetic tree was constructed using the Neighbor-joining method for analysing the *rrn* operon, 16S and 23S rRNA gene. Two clades produced in the phylogenetic trees were in agreement with the genomospecies A and B typing based on the amplification of the 23S rDNA region by
Istivan et al. (2004). The 16S rRNA and 23S rRNA gene sequences of eight *C. concisus* (seven of them were included in this study) have been analysed and used to construct phylogenetic trees (Deshpande et al., 2013). The phylogenetic tree based on the 23S rRNA gene sequences in the previous study is in agreement with the findings of our study when the whole *rrn* operon was used for the analysis. Interestingly, when a phylogenetic tree based on the 16S rRNA gene was constructed from the sequences used in this study, it was found that they were allocated into two clades based on the genomospecies which is in agreement with the *rrn* typing. The phylogenetic tree suggests that there may be some evolutionary development in *C. concisus* strains that has occurred through horizontal gene transfer and other mechanisms.

*C. concisus* is genetically diverse microorganism evidence provided by different studies using AFLP, RAPD or protein profiling. In this study, 12 *C. concisus* strains were successfully typed into two genomospecies (A & B) based on the indels identified in 5S rRNA, 23S rRNA, 16S rRNA genes and ITS regions within each operon, especially the sequences mentioned at indels 1, 2, 3, 12, 13 and 22. The phylogenetic tree constructed based on 15 *rrn* operons also supported the typing to genomospecies level. The 23S rRNA gene is suggested as the reliable region for *C. concisus* typing. This typing is consistent, reliable and explains the evolutionary development of *C. concisus*. 
Chapter Seven

General Discussion

_Campylobacter concisus_ is a hydrogen-requiring fastidious bacterium, first isolated in 1981 from a periodontal lesion (Tanner _et al._, 1981). Significant advances in the research of this bacterium have occurred over the last decades. Several studies reported _C. concisus_ as possible pathogen in the gastrointestinal tract (Vandamme _et al._, 1989; Lindblom _et al._, 1995; Istivan _et al._, 2004) while others reported _C. concisus_ to be associated with periodontal diseases (Macuch and Tanner, 2000; Moore _et al._, 1987; Kamma _et al._, 2001). Since 2009, it has been associated with chronic intestinal diseases (Zhang _et al._, 2009). But recently it has been established as a component of the normal microbiota in the human oral cavity (Petersen _et al._, 2007; Zhang _et al._, 2010; Zhang _et al._, 2014). Two or more genomospecies have been identified in this divergent species based on several typing methods including genomospecies A & B on the 23S rRNA gene amplification (Istivan _et al._, 2004). However, it is not confirmed yet if the oral strains are responsible for causing the previously mentioned diseases or if there are specific differences in the phenotypic and genotypic characters between the oral and intestinal strains.

Biofilm formation is considered as an important virulence factor. It protects the pathogen from the innate immunity of the host as well as supplies nutrient to the inhabitants of the biofilm (Jensen _et al._, 2010). Being a fastidious oral microorganism, biofilm formation is likely to be an important requirement for _C. concisus_ to survive in the oral cavity. However, there are only few studies on biofilm formation of this bacterium. The first report of _C. concisus_ capability in producing biofilms was by Gunther and Chen
Later Lavrencic et al. (2012) investigated the motility and biofilm formation of selected clinical *C. concisus* isolates and found all *C. concisus* were able to form biofilms. However, the study did not include any *C. concisus* isolated from the oral cavity. Until now, no thorough molecular study has been conducted on biofilm formation in *C. concisus*. The aim of this study was therefore to perform phenotypic investigation on the biofilm of both oral and intestinal *C. concisus*, and to investigate the effect of the signalling protein, LuxS, on biofilm formation. Comparative genomic analysis of selected oral and intestinal *C. concisus* was also performed to expand knowledge on potential virulence genes, and the genome structure of this species.

In this study, 14 intestinal and 19 oral *C. concisus* were included for investigation of the biofilm-forming ability. It was expected that all *C. concisus* tested strains would be able to form biofilms as this was previously reported (Gunther and Chen, 2009; Lavrencic et al., 2012). Interestingly, in this study it was found that the oral strains produced significantly higher levels of biofilm than the clinical isolates (*p* <0.05) and it was suggested biofilm in the oral cavity is an advantageous trait that enables to escape from toxic oxygen and other adverse conditions. In oral cavity mixed cultures in the biofilm can protect obligate anaerobes from the toxic effects of oxygen (Donlan and Costerton, 2002). This feature was also reported in *C. jejuni* NCTC 11168, where it formed biofilm more rapidly in aerobic conditions (20% O₂) than under microaerobic conditions (5% O₂, 10% CO₂) (Reuter et al., 2010). It is worthy to mention that RMIT-O17, the highest biofilm-forming strain belonged to genomospecies A and no significant association was found between biofilm formation and the genomospecies typing.
The highest biofilm-producing oral strain (RMIT-O17), and a moderate biofilm-producing intestinal strain, RCH 26, were selected for further investigation. Phenotypic characterization and developmental stages of RMIT-O17 biofilms were observed by phase contrast microscopy, confocal laser scanning microscopy (CLSM) and scanning electron microscopy (SEM). Biofilm development stages of RMIT-O17 were observed by phase contrast microscopy, and they were partitioned into the following three stages: attachment of planktonic cells to a surface; maturation of biofilm; and dispersion of single cells. These characteristic developmental biofilm stages are well established and have been observed in many other bacterial biofilms, including *Pseudomonas aeruginosa* (Sauer *et al*., 2002), *Streptococcus pneumonia* (Allegrucci *et al*., 2006) and *Salmonella Typhimurium* (Jonas *et al*., 2007).

Biofilm development of RMIT-O17 was investigated by CLSM which showed a complex matrix of live and dead cells in the mature biofilm. This is a feature of mature bacterial biofilms (Takenaka *et al*., 2001; Ica *et al*., 2012). The structure of the biofilm was also observed by the SEM technique and the images of RMIT-O17 biofilms revealed the presence of extracellular polymeric substance (EPS). It is another important characteristic of the biofilm which protects and supplies nutrients to the inhabitants in the biofilm (Flemming *et al*., 2016). The EPS has been identified in the biofilms of *C. jejuni* (Joshua *et al*., 2006), *P. aeruginosa* (Joo and Otto, 2012), and *Borrelia burgdorferi* (Sapi *et al*., 2012). Therefore, it can be concluded that biofilms produced by *C. concisus* may follow the similar configuration as other Gram negative bacteria.

Following the phenotypic characterization of *C. concisus* biofilms, a molecular investigation approach was undertaken and the *luxS* gene was investigated. LuxS is a
signalling protein responsible for quorum-sensing which has been linked to biofilm formation of many oral pathogens, such as *S. pneumoniae* (Vidal *et al.*, 2011), *S. mutans* (Merritt *et al.*, 2003), and *S. intermedius* (Ahmed *et al.*, 2009). LuxS was also correlated with biofilm formation in *C. jejuni* (Reeser *et al.*, 2007; Adler *et al.*, 2014), a closely related bacterium to *C. concisus*. The presence of luxS was confirmed by PCR in all intestinal and oral *C. concisus* isolates in this study. The phylogenetic tree generated from the available luxS gene within *C. concisus* sequences showed that the sequences were arranged into two clades in accordance with *C. concisus* genospecies typing (A and B) based on the 23S rRNA gene (Istivan *et al.*, 2004). However, there was no significant association between the biofilm formation capabilities and the genospecies typing, while the isolation site (oral or intestinal) was significantly associated with the amount of biofilms produced, which could be a survival mechanism as discussed above.

To further investigate the role of luxS in *C. concisus* biofilm formation, a luxS knock-out mutant was made in RMIT-O17 by inserting a kanamycin cassette. This is the first study to investigate the role of luxS in biofilm formation in *C. concisus*. The parental RMIT-O17 and the luxS-mutant (ΔluxS-O17) were compared for different phenotypic characteristics in biofilm formation and other virulence factors such as adhesion, invasion and motility. It was found that the quantity of biofilm formation was reduced significantly (in ΔluxS-O17, *p* <0.05), and that this reduction was also detected when biofilms were visualized by CLSM and SEM. An association between luxS (i.e. AI-2 signalling) and biofilm formation is established in many oral microorganisms (McNab *et al.*, 2003; Merritt *et al.*, 2003; Yoshida *et al.*, 2005) and other pathogens (Balestrino *et al.*, 2005; Cole *et al.*, 2004; Reeser *et al.*, 2007).
The motility of ΔluxS-O17 was significantly reduced ($p < 0.05$) when tested in the semi-solid agar medium. Being a signalling molecule, LuxS not only controls biofilm formation of bacteria, but it also influences motility by regulating flagellar biogenesis, as reported in *V. alginolyticus* (Tian *et al.*, 2008). Reduction in motility has been exhibited in *C. jejuni* after inactivation of LuxS (Elvers and Park, 2002; Holmes *et al.*, 2009; Adler *et al.*, 2014). LuxS product, AI-2 is thought to act as a global regulatory mechanism for basic physiological function and virulence factors (Adler *et al.*, 2014). However, to confirm the effect of LuxS in *C. concisus*, the restoration of motility should have been attempted by chemical or genetic complementation in ΔluxS-O17. Furthermore, RMIT-O17 has been found to be invasive in INT 407 with a very low invasion index (0.281) while the ΔluxS-O17 was rendered completely non-invasive in INT 407. In *C. jejuni* inactivation of luxS showed comparable invasion ability to Caco-2 cell monolayers to the parental strain (Elvers and Park, 2002). In another study, inactivation of luxS in *C. jejuni* 81-176 resulted in significantly reduced adherence in the LMH chicken hepatoma cells (Quinones *et al.*, 2009) and the adherence property could be restored by genetic complementation. It indicates the contribution of AI-2 in *C. jejuni* for colonization and survival ability in chicken. However, no significant change was found in the adhesion property of ΔluxS-O17. It suggests that LuxS in *C. concisus* may play a role not only in metabolic activity, such as motility, but also on virulence factors such as biofilm formation and invasion, but not on adhesion property.

In the second part of this thesis, four *C. concisus* stains were selected for genome sequencing and seek insight into genes related to virulence, and the genome structure of this diverse species. The strains selected were: the highly motile, biofilm-forming RMIT-O17 isolated from a healthy individual, another oral strain (RMIT-JF1) isolated from a
Crohn’s disease patient, a clinical intestinal isolate RCH 26 from a gastroenteritis patient, and a unique strain isolated from the duodenum of an IBD patient (AUS22-Bd2). To note, these four *C. concisus* strains (two oral and two intestinal) belonged to genomospecies A.

Data on four *C. concisus* genomes were produced by the next generation sequencers, Ion proton and Illumina Miseq. Although different assembly methods (Mira4, CLC genomic workbench, A5-miseq pipeline) were used, high N50 values indicated of good quality assemblies. Pairwise alignment using the Mauve alignment tool and comparative genomic circular view using Seed viewer based on the sequences of these four genomes with two reference genomes (*C. concisus* 13826 and ATCC 33237T) revealed interesting findings. Very high degrees of gene shuffling and inversion were exhibited when the four genomes were compared with *C. concisus* 13826; a strain belongs to genomospecies B, which is in agreement with a previous study (Deshpande et al., 2013). However, ATCC 33237T showed a very high level of similarity with the four *C. concisus* genomes sequenced in this study. It should be noted that these four *C. concisus* strains and the type strain ATCC 33237T belong to genomospecies A, and show a very high level of similarity.

Two plasmids were identified and characterized in the genome of RCH 26 only. Plasmid sequences were found after searched for the ‘ori’ region by the ‘Tandem Repeat Finder’ and confirmed by inverse PCR and Sanger sequencing. A unique plasmid was identified in RCH 26 (pCCON20) containing genes for the mobilization protein BmpH and the RepB (a replication protein). The high coverage number of pCCON20 indicated that it is a high copy number plasmid. The other plasmid (pCCON22) was larger than the first (pCCON20) and contained two contigs from the assembly. This plasmid shared eight
genes with *C. concisus* 13826 plasmid pCCON31 including genes encoded for a hemolysin activation/secretion protein and an outer membrane toxin activator protein CdiB-2. CdiB is an outer membrane pore-forming protein allowing secretion of the CdiA polymorphic toxin, a large filamentous protein (Aoki *et al.*, 2010). These are putative large exoproteins involved in heme utilization or adhesion of ShlA/HecA/FhaA family. In *C. concisus* the secreted haemolytic activity along with membrane-bound haemolytic phospholipase activity were reported by Istivan *et al.* (2008). Interestingly, neither of the two genomes of *C. concisus* oral strains, nor the ATCC 33237<sup>T</sup> reference genome contained any plasmid. The genes identified in pCCON20 were found to be homologous with genes from *C. coli*, *C. hominis* and Betaproteobacteria bacterium, which are all enteric pathogens. This indicates that *C. concisus* may acquire plasmids when present in the intestine mostly via horizontal transfer from other intestinal inhabitants.

Pan and core genome analysis revealed 1,463 core and 2,790 pan genes obtained from four *C. concisus* genomes sequenced in this study and from the two reference genomes. The core genome included genes in amino acid transport and metabolism, energy production and conversion and translation, ribosomal structure and biogenesis while the pan genome consisted of hypothetical genes and genes related to repair and replication. The Blast2GO analysis indicated that RNA processes, metabolic and biosynthetic processes were enriched in the core genome of *C. concisus* as expected. On the other hand, enrichment of defence responses, DNA-related processes such as DNA integration and DNA restriction-modification were detected in the pan genome, which supports the findings of Deshpande *et al.* (2013).
The orthologous clusters of genes in the four *C. concisus* genomes were also investigated. From this analysis, it was revealed that RCH 26, AUS22-Bd2, and RMIT-JF1 were sharing the most number of gene clusters. Virulence proteins or toxins such as M protein, FeMo cofactor biosynthesis protein NifB, prophage were shared among these *C. concisus* genomes and could be related with the virulence. Virulence associated genes were also investigated in individual genomes. RMIT-O17 being an adherent but non-EICC could be correlated with the presence of adhesion related proteins (NhaC and NhaA) and the absence of invasion related proteins (MobA, Exotoxin 9, StbE replicon stabilization toxin and RelB/StbD replicon stabilization protein). In addition to the absence of the previously mentioned proteins, another virulence related protein ‘Zot’ was not detected in the genome of RMIT-O17, which supports the hypothesis that oral strains from healthy individuals are more likely to be non-virulent. The *zot* was detected in both RCH 26 and RMIT-JF1, from a gastroenteritis patient and a Crohn’s disease patient respectively but not in the genome of AUS22-Bd2, an invasive strain isolated from the duodenum of an IBD patient. However, this invasive strain contained other virulence related genes in the genome, indicating that Zot may not be an obligatory virulence factor for *C. concisus*. It also indicated that virulence is not related to genetic typing of *C. concisus* to genomospecies level.

In Chapter 6, the first *in silico* analysis on the *rrn* operon (~6000 bp) of these four genomes and other eight available *C. concisus* genomes is reported. A total of 15 sequences were included in the analysis and 38 indels were identified in the *rrn* genomic loci in 12 *C. concisus* strains. The 23S rRNA gene was found to be diversified, yet possessed systematically arranged indels. Indels at several positions were significantly associated with either genomospecies A or B. Interestingly, the primers employed to
detect genospecies A and B by Istivan et al. (2004) were designed within indel 13 (122 bp long) which was identified in this study. In the 16S rRNA gene, 9 indels were identified, but they were not correlated to any genospecies. The ITS region was found highly variable in *C. concisus* with 3 indels and two tRNAs which is in agreement with previous study (Man et al., 2010b).

The phylogenetic tree constructed from these 15 *rrn* operons revealed interesting findings. In the phylogenetic trees constructed with the whole *rrn* operon, 23S rRNA gene and 16S rRNA gene sequences produced two clades which were in agreement with the previous typing system to genospecies A and B based on the amplification of the 23S rDNA region by Istivan et al. (2004). In a previous study the 16S rRNA genes from eight genomes were used to construct a phylogenetic tree and the strains isolated from patients with CD were grouped together (Deshpande et al., 2013). Yet in our study, seven of those eight sequences were included in the analysis and still the sequences were arranged into two clades of genospecies A and B based on 16S rRNA genes. The trees constructed on the basis of the *rrn* operon suggest the evolutionary relationship in *C. concisus*.

This study provides valuable information on *C. concisus* biofilm formation, its development and arrangement and phenotypic characteristics. The quorum-sensing gene LuxS was also shown to play a role in biofilm formation in *C. concisus*. In addition, the LuxS mutant also displayed reduced motility, invasion and adhesion, suggesting a central role in these processes in *C. concisus*. Genetic complementation is however essential to confirm these roles. *C. concisus* is a genetically diverse microorganism according to evidence provided by the whole genome sequencing. It was also known that the genetic diversity in *C. concisus* depends on the source of isolation. In this study, the differentiation
of *C. concisus* strains into two genomospecies (A & B) based on the indels identified in *rrn* operon was confirmed. The phylogenetic tree based on the 15 *rrn* operons also supported the typing to genomospecies level. To conclude, the results from this study indicate that the site of isolation rather than the genomospecies typing is more relevant to the biofilm-forming capabilities of the strains and to the presence of plasmids and virulence related properties. The phylogenetic tree arrangements based on the *rrn* operon were however more in alignment with the genomospecies typing rather than the site of isolation.
Future directions

In this study, biofilm formation by *C. concisus* was tested in polystyrene plates and in complex media; in order to better reflect the *in vivo* environment, it is proposed that that biofilm-forming ability should be tested on different natural nutrient sources like saliva and mucin. In this study, the phenotype of mature *C. concisus* biofilms were investigated and as a significant difference was observed between the biofilms of oral and intestinal *C. concisus*, a thorough investigation on the genes that are expressed during the formation of the biofilm should be undertaken to better understand the mechanism of biofilm formation by *C. concisus* on human teeth. Proteomic studies would be an ideal approach to investigate this. The ability of the intestinal isolates RMIT-JF1 and AUS22-Bd2 to form biofilms should be investigated to determine whether this feature relates to virulence. In addition the invasive capacity of RMIT-JF1 should also be tested and correlated with genetic findings.

Significant differences in biofilm formation, motility and invasion assay between RMIT-O17 and ΔluxS-O17 were observed in this study, and this needs to be confirmed by chemical or genetic complementation of luxS. Expression of genes involved in motility (*flaA, flaB* and *flaC*), adhesion, invasion and other putative virulence genes *cjaC, cjaA* and *dnaJ* in the wild type and ΔluxS-O17 should be investigated by semi-quantitative RT-PCR. Further studies may need to be done using animal models with *C. concisus luxS* mutant and the parental strain and to investigate and confirm the pathogenic role of LuxS.
References


References


References


Linton, D., Dewhirst, F. E., Clewley, J. P., Owen, R. J., Burnens, A. P. & Stanley, J. 1994. Two types of 16S rRNA gene are found in Campylobacter helveticus: analysis, applications and characterization of the intervening sequence found in some strains. Microbiology, 140 ( Pt 4), 847-55.


References


Figure A1: The PstI digest of phage-λDNA showing 29 fragments and size was used as the DNA marker.
Figure A2: Plasmid map of vector pBluescript II SK+
Appendix III

Figure A3: Plasmid map of vector pMW2
Appendix IV

Figure A4: Plasmid map of vector pBC SK+
Appendix V

>pCCON20
TTCGAGCGGCTCTAAAGGCTGGACTAGTACGAGCTAGCCTAGGATTTGTTAATCTTTAAGCGCAGCATATATACAAAACATCT
TGAGGCTGGAACGGCTGCTAAAGAAGGAGTAGATTGAAAGAAGGAGGACGAGAGTAGAAGGCTGAAAGAGAACAGT
CTAGAAAGACTGCTACAGCTAGAATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAA
ACATCAGCTAGAATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAG
AATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG
AGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACGAGAAGGAGTAG
ATTAGTACGAGAAGGAGTAGATTGAGTTTGTAAATCTTAACGGCACATATATAAAAACATCAGCTAGAATTAGTACG